UNIVERSIDAD DE SANTIAGO DE CHILE

FACULTAD DE CIENCIA

Departamento de Física

Regresión de datos en elipsometría de Mueller usando inteligencia artificial.

Simón Joaquín Paiva Ortega

Tesis para optar al Título Profesional de Ingeniero Físico.

Profesor Guía:

Felipe Herrera Urbina

Santiago – Chile

2020

© Simón Joaquín Paiva Ortega, 2020
Licencia Creative Commons Atribución-NoComercial Chile 3.0

“Regresión de datos en elipsometría de Mueller usando inteligencia artificial”

Simón Joaquín Paiva Ortega

Este trabajo de titulación fue preparado bajo la supervisión del (los) profesor(es) guía Dr. Felipe Herrera del Departamento de Física y ha sido aprobado por los siguientes miembros de la comisión calificadora del candidato:

(Firma)

…….

Dr. Dinesh Singh

(Firma)

…….

Dr. Francisco Vivanco

(Firma)

…….

Dr. Roberto Muñoz

(Firma)

…….

Dr. Roberto Bernal

Director Departamento de Física

Facultad de Ciencia

Resumen

Esta tesis es parte de un proyecto en el que se desarrolló un nuevo método para medir los índices de refracción y otras propiedades físicas de un cristal birrefringente. El método con el cual normalmente se estudian las propiedades ópticas de un cristal birrefringente es usando elipsometría para obtener distintos vectores se Stokes, los cuales al usar un método de reconstrucción pueden ser usados para obtener una matriz de Mueller, dicha matriz es separada en matrices de Mueller conocidas, esta separación sirve para poder ver cuál es el comportamiento del cristal y así tener una idea de sus propiedades físicas. Existen varios métodos de separación de los cuales ninguno es general, ósea ninguno cada uno sirve para un tipo de cristal, por lo que existen algoritmos de búsqueda que sirven para saber cuál es el método más conveniente al caracterizar materiales nuevos.

Este nuevo método se basa en la obtención de tres propiedades ópticas, las cuales son los dos índices de refracción del material (ordinario y extraordinario) y el ángulo del eje óptico con respecto al plano de incidencia del láser, nos basaremos en matrices de Mueller reconstruidas mediante los vectores de Stokes obtenidos experimentalmente, solo que para obtener las propiedades físicas se entrenara una Concatenate Neural Networks, la cual será entrenada con un algoritmo de trazado de rayos.

El algoritmo de trazados de rayos permite mediante el conocimiento de las propiedades físicas de un material y una polarización incidente, obtener la polarización de salida de la luz incidente y su orientación. Es por eso que usando este algoritmo para diferentes polarizaciones podemos obtener una matriz de Jones, la cual será usada para obtener una matriz de Mueller. Debido a esto es que debemos invertir el problema, ya que las matrices de Mueller pueden ser obtenidas mediante elipsometría y las propiedades ópticas del cristal son lo que deseamos obtener. Para invertir este problema usaremos Neuronal Networks lo que hace que el problema abarque cualquier tipo de cristal.

En esta tesis se estudiaron dos casos. El primero fue estudiar las propiedades ópticas de un cristal de cuarzo tipo α, cuya matriz de Mueller usada para obtener dichas propiedades es una matriz obtenida por una simulación usando el algoritmo de trazado de rayos para esto. El segundo fue estudiar un cristal de BBO, donde se vieron dos casos uno en el cual se obtuvo su matriz de Mueller para ser estudiada mediante el algoritmo de trazado de rayos y el segundo caso fue el usar una matriz obtenida experimentalmente mediante la reconstrucción usando vectores de Stokes para obtener las propiedades ópticas del cristal.

Palabras claves: cristal birrefringente, matriz de Mueller, Concatenate Neural Networks, Machine Learning, Artificial Neural Networks, trazado de rayos, polarimetría, elipsometría

Tabla de contenido

Introducción 1

Materiales Birrefringentes 1

Aplicaciones industriales 2

Caracterización de un cristal birrefringente usando elipsometría 4

Capítulo 1: Óptica de cristales birrefringentes 8

1.1 Ecuación de onda para materiales lineales 8

1.2 Ecuación de onda para materiales uniaxiales 9

1.3 Materiales uniaxiales 13

1.4 Ángulo de walk-off 15

1.5 Propagación de una onda polarizada a través de un material uniaxial 17

1.6 Algoritmo para obtener los coeficientes de Fresnel usando dos interfaces 22

1.6.1 Caso isotrópico-Birrefringente 22

1.6.2 Caso isotrópico-Birrefringente-isotrópico 24

Capítulo 2: Regresión de datos con aprendizaje de máquina supervisado 27

2.1 Aprendizaje de máquina (Machine Learning) 27

2.2 Artificial Neural Networks (ANN) 27

2.3 Feed-Forward Neural Networks 30

2.4 Machine Learning usando Python 32

2.5 Caso ejemplo: interpolación nolineal con una red neuronal 35

Capítulo 3: Elipsometría de Mueller usando redes neuronales pre-alimentadas (FFNN) 38

3.1 Elipsometría de Mueller 38

3.2 Elipsometría como problema de Machine Learning 39

3.3 Limitaciones del algoritmo 44

3.4 Caso de estudio A: cuarzo tipo α con matriz de Mueller simulada 46

3.5 Caso de estudio B: BBO usando una matriz experimental 51

Conclusiones 55

Referencias Bibliográficas 56

Índice de Tablas

Tabla 3.4.1 Datos predichos para el cristal de cuarzo tipo α usando matriz simulada **49**

Tabla 3.4.2 Datos predichos para el cristal de cuarzo tipo α usando matriz de componentes **49**

Tabla 3.5.1 Datos predichos para el cristal de BBO usando matriz simulada **51**

Tabla 3.5.2. Resultados para un cristal de BBO usando matriz experimental **52**

Tabla 3.5.3. Resultados para un cristal de BBO usando matriz de componentes ceros **53**

Índice de ilustraciones

Figura 0.1.1. Doble refracción o birrefringencia en un cristal de calcita **1**

Figura 0.1.2. Gráfico de la ecuación de Sellmeier $n(\lambda)$ para BBO tipo α **2**

Figura 0.2.1. Imagen de una Pockels cell basada en un cristal KDP **3**

Figura 0.2.2. Configuración típica del fenómeno de duplicación de frecuencias **3**

Figura 0.2.3. Configuración de un prisma Glan-Taylor. **4**

Figura 0.3.1. Esquema de medición experimental de elipsometría **6**

Figura 0.3.2. Esquema experimental más simple para un DOAP **7**

Figura 1.3.1. Tipos de polarización con respecto al plano de incidencia de los campos **13**

Figura 1.3.1. Muestra una onda incidente tipo ordinaria en un material uniaxial **15**

Figura 1.4.2. Variación entre el vector de poynting y el vector de onda dentro del BBO **16**

Figura 1.5.1. Refracción desde un material birrefringente hacia otro material birrefringente ...**17**

Figura 1.6.1. Rayo refractando desde un material isotrópico hacia un material uniaxial **21**

Figura 2.2.1 Ilustración de un Perceptron **27**

Figura 2.2.2 Ilustración de una ANN **28**

Figura 2.5.1 Comparación función coseno con datos predichos por una NN (0 a π radianes) **34**

Figura 2.5.2 Comparación función coseno con datos predichos por una NN (0 a 2π radianes) ..**35**

Figura 2.5.3. Gráfico del error cuadrático medio para la red neuronal del coseno **35**

Figura 3.2.1. Esquema general de un experimento de elipsometría de reflexión **41**

Figura 3.2.2 Esquema estándar de una concatenate neural network **41**

Figura 3.3.1 Variación de la intensidad con respecto al ángulo de incidencia para el cuarzo con polarización –s **44**

Figura 3.3.2 Variación de la intensidad con respecto al ángulo de incidencia para el cuarzo con polarización -p.....**44**

Figura 3.4.1. Esquema de la red neuronal usada para hacer la inversión del cuarzo tipo α.....**46**

Figura 3.4.2. Gráfico de la función de perdida “logcosh” **47**

Figura 3.4.3. Gráfico del error cuadrático medio de los datos de entrenamiento para la red neuronal del cuarzo **47**

Figura 3.4.4. Gráfico del error cuadrático medio de los datos de validación para la red neuronal del cuarzo (línea) **48**

Figura 3.4.4. Gráfico del error cuadrático medio de los datos de validación para la red neuronal del cuarzo (punto) **48**

Figura 3.5.1. Gráfico del error cuadrático medio de los datos de entrenamiento para la red neuronal del BBO **50**

Figura 3.5.2. Gráfico del error cuadrático medio de los datos de validación para la red neuronal del BBO **51**

Introducción

1) Materiales Birrefringentes:

En cierta clase de cristales transparentes, la velocidad de fase de la luz depende de la polarización del haz incidente. Este tipo de materiales son llamados birrefringentes o materiales doble refractantes (Jhon F. Ready, 1978). Este fenómeno se debe a que dentro del material existen dos índices de refracción, más específicamente, existe un eje dentro del cristal que tiene un índice de refracción distinto a los demás ejes. Este es denominado eje extraordinario y el índice de refracción que pasa por este eje también es llamado índice extraordinario. Así mismo, los otros son denominados ejes ordinarios e índices ordinarios. Donde el cristal birrefringente más usado en aplicaciones es el Cuarzo (G. Q. Zhou, 1998) debido a su gran abundancia natural. Un ejemplo de birrefringencia es la calcita como se puede ver en la figura 0.1.1.

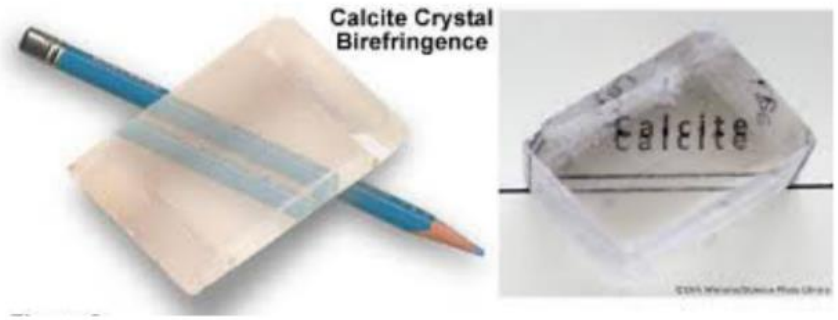

Figura 0.1.1. Doble refracción o birrefringencia en un cristal de calcita.

Una forma de caracterizar a los cristales uniaxiales o birrefringentes es midiendo los dos índices de refracción que tiene. Debido a que dentro de un material el índice de refracción varía dependiendo de la longitud de onda incidente, se puede establecer una relación empírica llamada ecuación de Sellmeier (W. Sellmeier, 1871) de siguiente forma:

$$n^2(\lambda) = 1 + \frac{B_1\lambda^2}{\lambda^2 - C_1} + \frac{B_2\lambda^2}{\lambda^2 - C_2} + \frac{B_3\lambda^2}{\lambda^2 - C_3} \qquad (0.1.1)$$

Donde las constantes B_1, B_2, B_3, C_1, C_2 y C_3 son únicas para cada material, y se determinan experimentalmente. Para el BBO tipo α las curvas de Sellmeier se muestran en la siguiente figura 0.1.1, el cual es un material birrefringente.

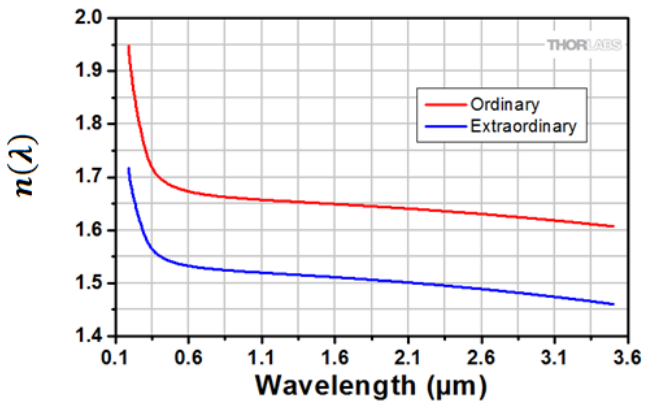

Figura 0.1.2. Gráfico de la ecuación de Sellmeier $n(\lambda)$ para un cristal birrefringente BBO tipo α desde los 190 nm hasta los 3.5 µm (Appel, R,2002).

Lo que corresponde a las ecuaciones de Sellmeier (G. Tamosauskas, 2018):

$$n_o^2(\lambda) = 1 + \frac{0.90291\,\lambda^2}{\lambda^2 - 0.003926} + \frac{0.83155\,\lambda^2}{\lambda^2 - 0.018786} + \frac{0.76536\,\lambda^2}{\lambda^2 - 60.01} \qquad (0.1.2)$$

$$n_e^2(\lambda) = 1 + \frac{1.151075\,\lambda^2}{\lambda^2 - 0.007142} + \frac{0.21803\,\lambda^2}{\lambda^2 - 0.02259} + \frac{0.656\,\lambda^2}{\lambda^2 - 263} \qquad (0.1.3)$$

La forma de obtener estas ecuaciones experimentalmente es a través de elipsometría (X. Wan, 2004), este proceso se basa en el estudio de las polarizaciones incidentes y de salida de un material, con las cuales se pueden obtener una matriz de transferencia, la cual es un resumen de como el material afecta a las polarizaciones incidentes. Esta caracterización va a ser descrita de mejor forma en partes posteriores de esta tesis.

2) Aplicaciones industriales.

Los materiales birrefringentes tienen varias aplicaciones industriales, dentro de las cuales están los Pockels Cells (R, Paschotta, 2008),switches ópticos(R, Paschotta, 2008), cristales líquidos(R, Paschotta, 2008), modulador optoacustico(R, Paschotta, 2008), SHG (W. Boyd, 2003), polarizadores(R. Ulrich ,1980), entre otros.

Una Pockels Cell es un cristal electro-optico por el cual propaga un haz de luz. Que experimenta un retraso en la fase (efecto Pockels). Esta fase puede ser modulada aplicando un voltaje eléctrico variable. La Pockels cell actúa como un controlador de voltaje tipo waveplate, y es el componente básico de los moduladores ópticos usados en los lasers pulsados Q swicthing. Los materiales comúnmente usados como Pockels cells son el KDP (potassium di-deuterium phosphate), KTP (potassium titanyl phosphate), BBO (β-barium borate),

$LiNbO_3$(lithium niobate), $LiTaO_3$(lithium tantalate) y ADP ($NH_4H_2PO_4$, ammonium dihydrogen phosphate). Todos estos materiales son cristales no lineales. El KDP, KTP y BBO son materiales uniaxiales o birrenfrigentes(R, Paschotta, 2008). En la figura 0.2.1 podemos ver una imagen de una Pocket Cell basada en un cristal de KDP.

Figura 0.2.1. Imagen de una Pockels cell basada en un cristal birrefringente KDP, esta Pockels cell puede ser usada como base para un Q switching de un láser de estado sólido (R, Paschotta, 2008).

Los moduladores de cristales líquidos, son moduladores ópticos que se ocupan generalmente en las pantallas de cristal líquido (LCD por sus siglas en ingles), también pueden ser usados para modular un haz de luz láser.

Los cristales que no tienen centro de simetría pueden tener no linealidades de segundo orden. Esta no linealidad puede llevar a cabo el proceso de doblado de frecuencias, como se muestra en la configuración de la figura 0.2.2, donde a partir de un input se genera otra onda que posee el doble de la frecuencia que la onda incidente. Este fenómeno se usa principalmente para generar pulsos ultracortos y los cristales más usados para estas tareas son: KDP, KTP, BBO y LBO (P. A. Franken, 1961)(A. Ashkin, 1966).

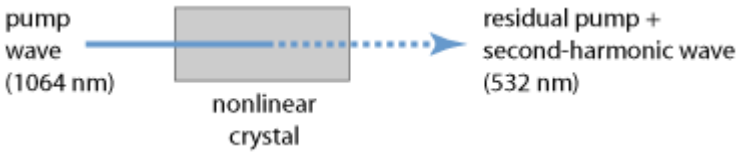

Figura 0.2.2. Configuración típica del fenómeno de duplicación de frecuencias: donde de input tenemos un rayo infrarojo de 1064 [nm], el cual pasa por un cristal no lineal, lo que genera un rayo de salida verde de 532 [nm] (R, Paschotta, 2008).

Otras de las aplicaciones industriales más comunes para los materiales birrefringentes es la construcción de polarizadores tipo beam-spliters, los cuales separan un haz en dos y según las propiedades internas del material pueden tener distintos efectos en un rayo del láser. Algunos de estos cristales pueden polarizar de forma completa los dos rayos de salida, los cuales son llamados prismas de Wollaston. En otros polarizadores con cristales birrefringentes ocurre una reflexión total interna para cierto tipo de polarización incidente, obteniendo

diferente dirección de los rayos para diferentes polarizaciones, como se ilustra en la figura 0.2.3.

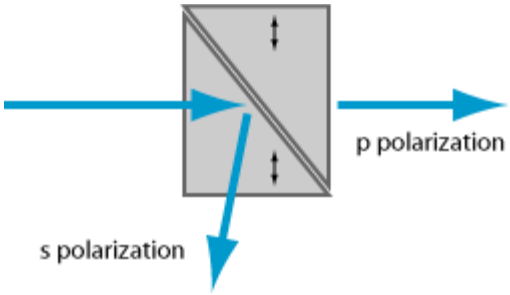

Figura 0.2.3. Configuración de un prisma Glan-Taylor. Se produce una reflexión total interna para la polarización tipo s, cambiando la dirección de salida de este tipo de polarización, pero no afecta la dirección del estado de polarización tipo p (R, Paschotta, 2008).

Otra aplicación industrial de los materiales birrefringentes es la fabricación de lasers, con el propósito de que la salida del láser sea linealmente polarizada sin tener pérdidas por depolarización dentro del cristal. Posteriormente veremos que las aproximaciones hechas para esta tesis son justamente para procesos sin depolarizaciones o con depolarizaciones muy pequeñas (R. Ulrich ,1980).

3) Caracterización de un cristal birrefringente usando elipsometría.

Consideremos una onda plana monocromática con frecuencia v viajando en dirección z con velocidad c. El campo eléctrico viajara en el plano *x-y*, y está descrito por:

$$\vec{E}(z,t) = Re\left(A \cdot e^{\left[i2\pi v\left(t-\frac{z}{c}\right)\right]}\right), \tag{0.3.1}$$

dónde:

$$\vec{A} = A_x\hat{x} + A_y\hat{y} \tag{0.3.2}$$

Es un vector con componentes complejos A_x y A_y. Para describir la polarización de esta onda trazamos el punto final del vector $\vec{E}(z,t)$ en cada posición de z como una función del tiempo.

Usando la ecuación (3.1.1) podemos caracterizar completamente una onda plana monocromática por las siguientes envolventes complejas, para describir los componentes *x* e *y* del campo electromagnético:

$$A_x = a_x e^{i\varphi_x} \tag{0.3.3}$$

$$A_y = a_y e^{i\varphi_y} \tag{0.3.4}$$

Ahora si escribimos estas componentes en forma de una columna matricial, como los vectores de Jones (M. Bass, 1995).

$$\boldsymbol{J} = \begin{bmatrix} A_x \\ A_y \end{bmatrix}, \tag{0.3.5}$$

Podemos calcular la intensidad de la onda usando:

$$I = \frac{|A_x|^2 + |A_y|^2}{2\eta} \tag{0.3.6}$$

Donde η corresponde a la impedancia del medio (M. Bass, 1995). Para una la intensidad normalizada tenemos:

$$1 = |A_x|^2 + |A_y|^2 \tag{0.3.7}$$

Ahora supongamos que una onda con las características mencionadas con anterioridad pasa a través de un material, debe haber una matriz de transición que relacione el vector de Jones de entrada con el de salida. De la siguiente forma

$$\boldsymbol{J_2} = T\boldsymbol{J_1} \tag{0.3.8}$$

Donde

$$T = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \tag{0.3.9}$$

Es la matriz de Jones.

Con los elementos introducidos con anterioridad tenemos una base teórica mínima para entender de mejor forma cómo podemos estudiar elipsometría de forma experimental. Existen varios métodos para calcular ya sea la matriz de jones como la matriz relacionada de Mueller (la cual va a ser introducida posteriormente en esta tesis) que permiten caracterizar un material óptico dependiendo de la polarización de la onda incidente.

En la figura 0.3.1, se muestra un típico esquema de medición para hacer polarimetría experimental

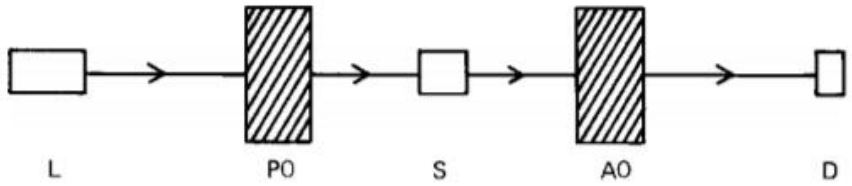

Figura 0.3.1. Esquema de medición experimental genérico usado para hacer elipsometría donde L representa la fuente de luz, PO es un polarizador óptico, AO es un analizador óptico, S es la muestra que quiere ser medida y D corresponde a un fotodetector (M. Bass, 1995. Capítulo 27, elipsometría).

Este esquema consiste: primero en polarizar un haz de luz con el polarizador PO, en alguna de las polarizaciones estándar como horizontal o vertical. Con esto se construye un vector de jones $\boldsymbol{J_1}$ según la ecuación (3.1.9). Después el haz pasa por la muestra cambiando las polarizaciones en el haz de salida. Con el analizador óptico se calculan las polarizaciones de salida y con el fotodetector se calcula la intensidad de salida en el punto D de la figura (0.3.1), obteniendo con esto un vector de Jones $\boldsymbol{J_2}$. Esto se hace para poder comparar la intensidad de salida con la de entrada en el punto L de la misma figura, estos cálculos pueden ser útil para calcular el grado de depolarizacion del cristal para cierta polarización incidente, como en el caso de los prismas de Wollaston.

Otro esquema de medición experimental usado es el DOAP (division-of-amplitude photopolarimeter), en el cual el arreglo más simple consiste en uno de cuatro fotodetectores de estado sólido y ningún otro elemento óptico, como se muestra en el esquema de la figura 0.3.2. Los primeros tres fotodetectores reflejan solo parcialmente y especularmente, y el cuarto (el ultimo) tiene un recubrimiento antirreflejo, lo que hace que esta última interface sea casi totalmente absorbente. En este tipo de DOAP, las cuatro señales de salida de los cuatro fotodetectores definen el vector $I = [I_0, I_1, I_2, I_3]^t$, la cual esta linealmente relacionado con un nuevo vector que llamaremos S de la siguiente forma:

$$I = AS \tag{0.3.10}$$

Donde $S = [S_0, S_1, S_2, S_3]^t$, y es denominado el vector de Stokes (el cual será introducido de mejor forma más adelante en esta tesis), en este caso este vector es el vector de Stokes incidente y el subíndice t indica que es la transpuesta del vector. La matriz instrumental A de 4x4 es determinada mediante calibraciones (R . M . A . Azzam, 1989) usando un polarizador lineal junto con un retardador de cuarto de onda. Una vez que esta matriz es obtenida podemos determinar el vector de Stokes de la señal de salida de la siguiente forma

$$S = A^{-1}I \tag{0.3.11}$$

Donde A^{-1} corresponde a la matriz inversa de A (M. Bass, 1995).

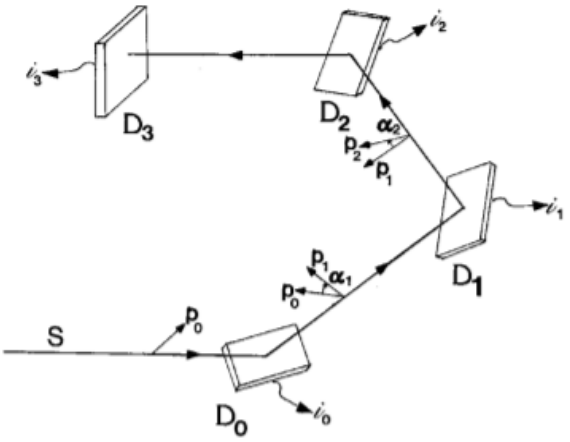

Figura 0.3.2. Esquema experimental más simple para un DOAP (division-of-amplitude photopolarimeter) donde los primeros tres fotodetectores (D_0, D_1, D_2) son parcialmente reflectantes y el ultimo es casi totalmente reflectante (M. Bass, 1995).

Capítulo 1: Óptica de cristales birrefringentes

1.1 Ecuación de onda para materiales lineales

La teoría fundamental de los campos electromagnéticos está basada en las ecuaciones de Maxwell (Partha, 2003) (Jackson, 1999). Que en su forma diferencial están expresadas de la siguiente forma

$$\vec{\nabla} \cdot \vec{\mathrm{D}} = \rho \tag{1.1.1}$$

$$\vec{\nabla} \cdot \vec{\mathrm{B}} = 0 \tag{1.1.2}$$

$$\vec{\nabla} \times \vec{\mathrm{E}} = -\frac{\partial \vec{\mathrm{B}}}{\partial \mathrm{t}} \tag{1.1.3}$$

$$\vec{\nabla} \times \vec{\mathrm{H}} = \vec{J_c} + \frac{\partial \vec{\mathrm{D}}}{\partial \mathrm{t}} \tag{1.1.4}$$

donde $\vec{E}$ es el campo eléctrico, $\vec{H}$ es el campo magnético, $\vec{D}$ es la densidad de flujo eléctrico, $\vec{B}$ es la densidad de flujo magnético, $\vec{J}$ es la densidad de corriente externa y ρ es la densidad de carga eléctrica externa. La polarización inducida depende linealmente del campo eléctrico según la ecuación (Boyd, 2003).

$$\vec{P}(\omega) = \chi^{(1)}\, \vec{E}(\omega) (Ec. 1.1.5)$$

Donde $\chi^{(1)}$ es conocida como la susceptibilidad lineal, $P(t)$ en la polarización y $E(t)$ el campo aplicado al material. Además tenemos la siguiente relación entre los campos y las densidades de flujo si es que la fuente externa está en el vacío (Jackson, 1999)

$$\vec{D} = \varepsilon_0 \vec{E} \tag{1.1.6}$$

$$\vec{B} = \mu_0 \vec{H} \tag{1.1.7}$$

Tomando el Rotor de la ecuación (1.1.3) y usando la ecuación (1.1.7) obtenemos

$$\vec{\nabla} \times \vec{\nabla} \times \vec{\mathrm{E}} = -\vec{\nabla} \times \frac{\partial \vec{\mathrm{B}}}{\partial \mathrm{t}} = -\frac{\partial}{\partial \mathrm{t}} \vec{\nabla} \times \vec{\mathrm{B}} = -\mu_0 \frac{\partial}{\partial \mathrm{t}} \left(\vec{\nabla} \times \vec{\mathrm{H}}\right) \tag{1.1.8}$$

Si usamos la ecuación (1.1.4) para substituir el rotor de $\vec{H}$ obtenemos y la ecuación (1.1.6), obtenemos

$$\vec{\nabla} \times \vec{\nabla} \times \vec{\mathrm{E}} = -\mu_0 \varepsilon_0 \frac{\partial^2 \vec{\mathrm{E}}}{\partial \mathrm{t}^2} - \mu_0 \frac{\partial \vec{\mathrm{J}}}{\partial \mathrm{t}} \tag{1.1.9}$$

Reemplazando la siguiente relación vectorial

$$\vec{\nabla} \times \vec{\nabla} \times \vec{A} = \vec{\nabla}(\vec{\nabla} \cdot \vec{A}) - (\vec{\nabla} \cdot \vec{\nabla})\vec{A} \tag{1.1.10}$$

En la ecuación (1.1.9), obtenemos

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - (\vec{\nabla} \cdot \vec{\nabla})\vec{E} = -\mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} - \mu_0 \frac{\partial \vec{J}}{\partial t} \tag{1.1.11}$$

Y finalmente reemplazando la ecuación 1.1.1 y la ecuación 1.1.6 en la ecuación 1.1.1, obtenemos la ecuación de onda

$$\nabla^2 \vec{E} = \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} + \mu_0 \frac{\partial \vec{J}}{\partial t} - \frac{1}{\varepsilon_0} \vec{\nabla} \rho \tag{1.1.12}$$

El espacio está libre de toda fuente, es decir

$$\vec{J} = 0 \tag{1.1.13}$$

$$\rho = 0 \tag{1.1.14}$$

La ecuación (1.1.12) queda como

$$\nabla^2 \vec{E} - \mu_0 \varepsilon_0 \frac{\partial^2 \vec{E}}{\partial t^2} = 0 \tag{1.1.15}$$

Que corresponde a la ecuación de onda homogénea para un sistema libre de fuente en el vacío.

1.2 Ecuación de onda para materiales uniaxiales.

Para el caso específico de los materiales uniaxiales, estudiados en esta tesis, necesitamos obtener una ecuación de onda libre de fuente parecida a la obtenida en la sección 1.1. En este caso no usaremos la expresión usada en 1.1.5 como una ecuación para la polarización inducida por un campo aplicado, sino que la siguiente expansión en serie de la polarización (Boyd, 2003)

$$\vec{P}(t) = \chi^{(1)}\, \vec{E}(t) + \chi^{(2)}\, \vec{E}^2(t) + \chi^{(3)}\, \vec{E}^3(t) + \cdots \tag{1.2.1}$$

donde los términos $\chi^{(2)}\ y\ \chi^{(3)}$ corresponden a la susceptibilidad de segundo y tercer orden respectivamente. Nos referiremos a una polarización de segundo orden dada por

$$\vec{P}^{(2)} = X^{(2)}\, \vec{E}^2(t) \tag{1.2.2}$$

Y a la polarización de tercer orden como:

$$\vec{P}^{(3)} = X^{(3)}\, \vec{E}^3(t) \tag{1.2.3}$$

Cabe mencionar que estas ecuaciones solo son válidas si es que el material no presenta perdidas ni dispersión.

Ahora para obtener la ecuación de onda de un material no lineal, primero asumimos que el material es no magnético, por lo que

$$\vec{H} = \vec{B} \tag{1.2.4}$$

Sin embargo, permitimos que el material sea no lineal relacionando a $\vec{D}$ con $\vec{E}$ de la siguiente forma (Boyd, 2003):

$$\vec{D} = \vec{E} + 4\pi\vec{P} \tag{1.2.5}$$

Y usando las ecuaciones de Maxwell de la sección anterior, con $\rho = J_c = 0$ obtenemos la siguiente ecuación

$$\vec{\nabla} \times \vec{\nabla} \times \vec{\mathrm{E}} = -\mu_0\varepsilon_0 \frac{\partial^2 \vec{\mathrm{D}}}{\partial \mathrm{t}^2} \tag{1.2.6}$$

Substituyendo para $\vec{\mathrm{D}}$ la ecuación anterior y usando la ecuación (1.2.5) obtenemos

$$\vec{\nabla} \times \vec{\nabla} \times \vec{\mathrm{E}} = -\mu_0\varepsilon_0 \frac{\partial^2 \vec{\mathrm{E}}}{\partial \mathrm{t}^2} - 4\pi\mu_0\varepsilon_0 \frac{\partial^2 \vec{\mathrm{P}}}{\partial \mathrm{t}^2} \tag{1.2.7}$$

Usando la relación de la ecuación $(Ec. 1.1.10)$ y las relaciones para un medio sin fuente, obtenemos

$$-\vec{\nabla}^2 \vec{\mathrm{E}} = -\mu_0\varepsilon_0 \frac{\partial^2 \vec{\mathrm{E}}}{\partial \mathrm{t}^2} - 4\pi\mu_0\varepsilon_0 \frac{\partial^2 \vec{\mathrm{P}}}{\partial \mathrm{t}^2} \tag{1.2.8}$$

A continuación, expandimos $\vec{P}$ de la siguiente forma:

$$\vec{\mathrm{P}} = \vec{\mathrm{P}}^{(1)} + \vec{\mathrm{P}}^{\,NL} \tag{1.2.9}$$

donde $\vec{\mathrm{P}}^{(1)}$ es la parte de la polarización que depende linealmente del campo incidente $\tilde{E}$, así mismo podemos descomponer el campo $\tilde{D}$ en una parte lineal y una parte no lineal de la forma:

$$\vec{\mathrm{D}} = \vec{\mathrm{D}}^{(1)} + 4\pi\vec{\mathrm{P}}^{NL} \tag{1.2.10}$$

Donde la parte lineal de $\tilde{D}$ esta dado por

$$\vec{\mathrm{D}}^{\,(1)} = \vec{E} + 4\pi\vec{\mathrm{P}}^{(1)} \tag{1.2.11}$$

Usando las ecuaciones (1.2.9), (1.2.10) y (1.2.11) en la ecuación (1.2.8) obtenemos la siguiente ecuación de onda

$$-\nabla^2\vec{E} + \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\vec{\mathrm{D}}^{(1)} = -\frac{4\pi}{c^2}\frac{\partial^2 \vec{\mathrm{P}}^{NL}}{\partial t^2} \tag{1.2.12}$$

En este si es que la parte derecha de la ecuación fuera igual a cero, esta ecuación sería una ecuación de onda homogénea. Pero debido a que es distinta de cero esta parte de la ecuación actúa como fuente de radiación.

Ahora consideremos un medio sin pérdidas ni dispersión, por lo tanto, podemos hacer una relación entre los campos $\widetilde{D}^{(1)}$ y $\tilde{E}$ en términos de un tensor dieléctrico real e independiente de la frecuencia $\epsilon^{(1)}$, dada por

$$\vec{\mathrm{D}}^{(1)} = \epsilon^{(1)} \cdot \vec{E} \tag{1.2.13}$$

donde, si usamos el sistema de coordenadas de ejes principales *x, y, z*

$$\epsilon^{(1)} = \begin{bmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{bmatrix} \tag{1.2.14}$$

Para el caso de materiales isotrópicos, es decir, donde las propiedades del material no dependen de la dirección de incidencia del campo, la ecuación 1.2.13 se reduce a

$$\vec{\mathrm{D}}^{(1)} = \epsilon^{(1)} \vec{E} \tag{1.2.15}$$

Donde $\epsilon^{(1)}$ pasa de un tensor a una cantidad escalar. La ecuación de la onda obtenida en la ecuación (1.2.12) para un material isotrópico y sin dispersión queda como

$$-\nabla^2 \vec{E} + \frac{\epsilon^{(1)}}{c^2} \frac{\partial^2}{\partial t^2} \vec{E}^{(1)} = -\frac{4\pi}{c^2} \frac{\partial^2 \vec{\mathrm{P}}^{NL}}{\partial t^2} \tag{1.2.16}$$

Esta ecuación tiene la forma la ecuación de onda inhomogénea, donde la respuesta no lineal del material actúa como fuente.

Para el caso de un medio dispersivo. Esto quiere decir que se cumple la siguiente relación

$$\frac{d\, n(\omega)}{d\, \omega} \neq 0 \tag{1.2.17}$$

debemos considerar todas las frecuencias que componen al campo de forma separada de la siguiente forma

$$\vec{E}_n(r,t) = \sum_n \vec{E}_n(r) e^{-i\omega_n t} + c.c \tag{1.2.18}$$

$$\vec{D}_n^{(1)}(r,t) = \sum_n \vec{D}_n^{(1)}(r) e^{-i\omega_n t} + c.c \tag{1.2.19}$$

$$\vec{P}_n^{NL}(r,t) = \sum_n \vec{P}_n^{NL}(r) e^{-i\omega_n t} + c.c \tag{1.2.20}$$

Si ahora generalizamos el caso en el cual el medio presenta disipación, esto permite que el tensor dieléctrico que relaciona las amplitudes complejas de los campos de la siguiente forma

$$\vec{D}_n^{(1)}(\omega) = \epsilon^{(1)}(\omega_n) \cdot \vec{E}_n(\omega) \tag{1.2.21}$$

Con la expresión de la ecuación 1.2.18 podemos reescribir la ecuación de onda para cada modo de la forma

$$-\nabla^2 \vec{E}_n - \frac{\omega_n^2}{c^2} \epsilon^{(1)}(\omega_n) \cdot \vec{E}_n(\omega) = \frac{4\pi\omega_n^2}{c^2} \vec{P}_n^{NL}(\omega) \tag{1.2.22}$$

Para el caso de un material uniaxial o birrefringente solo habría que substituir la polarización no lineal por la polarización de segundo orden de la ecuación (1.2.2), obteniendo

$$-\nabla^2 \vec{E}_n - \frac{\omega_n^2}{c^2} \epsilon^{(1)}(\omega_n) \cdot \vec{E}_n(\omega) = \frac{4\pi\omega_n^2}{c^2} \chi^{(2)} \vec{E}_n^{\,2}(\omega) \tag{1.2.23}$$

Expandiendo la ecuación 1.2.2 en términos de *i, j, k* y reemplazando obtenemos

$$-\nabla^2 \vec{E}_i - \frac{\omega_i^2}{c^2} \epsilon^{(1)}(\omega_i) \cdot \vec{E}_i(\omega) = \frac{4\pi\omega_i^2}{c^2} \chi_{ijk}^{(2)} \vec{E}_j \vec{E}_k(\omega) \tag{1.2.24}$$

1.3 Materiales uniaxiales.

En esta tesis hablaremos de dos tipos de polarización distintas, definidas según el plano de incidencia del campo electromagnético. Estas dos polarizaciones son la polarización –s que tiene una incidencia del campo electromagnético paralelo con el plano de incidencia al material y una polarización –p la cual implica un campo de incidencia perpendicular al material (Robin H. A. Ras, 2007) como se puede ver en la figura 1.3.1. Este plano de incidencia esta descrito por el vector de incidencia $\vec{k}_i$ y la normal $\vec{\eta}$.

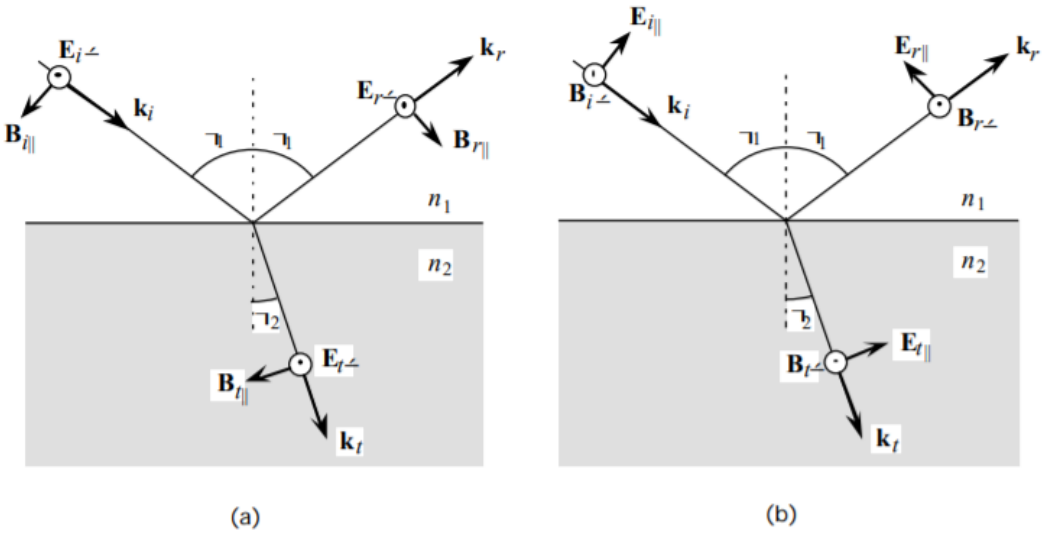

Figura 1.3.1. Tipos de polarización con respecto al plano de incidencia de los campos (a). Polarización incidente tipo –p (perpendicular) al plano del incidencia del material con índice de refracción n_2 (b). Polarización incidente tipo –s (paralelo) al plano de incidencia al material con índice de refracción n2 (A. Beléndez, 1996) (Polarización).

Según se mencionó en la introducción, los materiales uniaxiales poseen dos índices de refracción distintos por lo que existen 2 tipos de materiales uniaxiales: el tipo positivo en el cual el índice de refracción extraordinario es mayor que el índice de refracción ordinario ($n_e > n_o$) y el negativo que es el caso contrario ($n_0 > n_e$) (Boyd, 2003).

Para determinar el índice de refracción real que siente una onda propagando por una dirección arbitraria dentro de un cristal uniaxial, primero debemos determinar los modos normales para una onda plana viajando en una dirección arbitraria $\hat{u}$. Para esto usamos las ecuaciones de Maxwell (1.1.1) a (1.1.4) y la relación constitutiva (Bahaa E. A. Saleh, 1991):

$$D = \epsilon E \tag{1.3.1}$$

donde

$$\epsilon = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} \tag{1.3.2}$$

Se asume que todos los campos varían con la posición $\vec{r}$ de la forma $e^{-i(\vec{k}\cdot\vec{r})}$, donde $\vec{k} = k\hat{u}$. Las ecuaciones de maxwell (1.1.3) y (1.1.4) quedan de la siguiente forma

$$\vec{k} \times \vec{H} = -\omega\vec{D} \tag{1.3.3}$$

$$\vec{k} \times \vec{E} = -\omega\mu_o\vec{H} \tag{1.3.4}$$

Ahora substituimos (1.3.4) en (1.3.3) y usando (1.3.1) obtenemos:

$$\vec{k} \times \left(\vec{k} \times \vec{E}\right) + \omega^2\mu_o\,\epsilon\vec{E} = 0 \tag{1.3.5}$$

Esta ecuación vectorial puede ser vista en base a las tres componentes del campo eléctrico $\vec{E}_1, \vec{E}_2, \vec{E}_3$, con respecto a los tres ejes principales del cristal, obteniendo la siguiente matriz:

$$\begin{bmatrix} n_1^2k_0^2 - k_2^2 - k_3^2 & k_1k_2 & k_1k_3 \\ k_2k_1 & n_2^2k_0^2 - k_1^2 - k_3^2 & k_2k_3 \\ k_3k_1 & k_3k_2 & n_3^2k_0^2 - k_1^2 - k_2^2 \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{1.3.6}$$

Donde (k_1, k_2, k_3) son las componentes de k, $k_0 = \omega\sqrt{\mu_0\epsilon_0}$ y (n_1, n_2, n_3) corresponden a los índices de refracción en cada eje del cristal, para nuestro caso específico en el cual se están viendo materiales uniaxiales $n_1 = n_2 = n_o$ y $n_3 = n_e$. Usando esto y calculando el determinante de la expresión de la izquierda obtenemos:

$$(k^2 - n_o^2k_0^2)\left(\frac{k_1^2}{n_1} + \frac{k_2^2}{n_2} + \frac{k_3^2}{n_3^2} - k_0^2\right) = 0 \tag{1.3.7}$$

Existen dos posibles soluciones para la ecuación (1.3.7), la primera describe una esfera y tiene la siguiente forma:

$$k^2 = n_o^2k_0^2 \tag{1.3.8}$$

Y la segunda describe un elipsoide de revolución, tiene la siguiente forma:

$$\frac{k_1^2}{n_1} + \frac{k_2^2}{n_2} + \frac{k_3^2}{n_3^2} = k_0^2 \tag{1.3.9}$$

Donde la primera solución es la solución para la onda viajando por un plano en el cual no se encuentra el eje óptico, esto implica que el material actúa como un material isotrópico cualquiera ($n_1 = n_2 = n_3$) y la segunda solución es la solución para una onda viajando dentro del plano del eje óptico siguiendo la figura 1.3.1

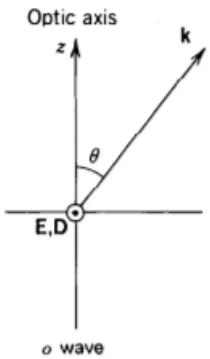

Figura 1.3.1. Muestra una onda incidente tipo ordinaria, en un material uniaxial con el eje óptico en el eje z y un ángulo de incidencia θ (Bahaa E. A. Saleh, 1991) (Capitulo 6).

Ahora supongamos que el eje donde se encuentra el índice de refracción extraordinario es el eje z, como lo indica la figura 1.3.1, entonces si es que $k_3 = k_z$ y usando como referencia la figura z podemos reemplazar $k_z = \cos(\theta)$, de la misma forma si hacemos $k_1 = k_x$ y $k_2 = k_y$. Como el plano de incidencia con respecto a los ejes *x* e *y* está definido de manera arbitraria tenemos que $\sqrt{k_1^2 + k_2^2} = \sin(\theta)$. Además sabiendo que k_0 corresponde al k incidente en el medio obtenemos que el índice de refracción de la onda $n^2(\theta)$ esta dado por

$$\frac{\sin(\theta)^2}{n_e^2} + \frac{\cos(\theta)^2}{n_o^2} = \frac{1}{n^2(\theta)} \tag{1.3.10}$$

1.4 Ángulo de walk-off

Sabemos que cuando una onda electromagnética pasa por un cristal uniaxial, si es que el plano de incidencia contiene al eje óptico o eje extraordinario como se muestra en la figura (1.3.1) y si es que existe un ángulo θ con respecto a este eje distinto a $\frac{\pi}{2}$ entonces la onda "sentirá" dos índices de refracción distintos por lo que se bifurcaría en una onda llamada onda extraordinaria (siente el índice de refracción dado por la ecuación (1.3.9)) y otra llamada onda ordinaria (siente el índice ordinario n_o), por lo tanto la onda transmitida dentro del material y la energía transmitida (vector de poynting) tienen distintas direcciones. El ángulo que separa la onda transmitida con la energía transmitida es llamado el ángulo de *walk-off* o ángulo de birrefringencia en el caso de materiales uniaxiales (V.G. Dmitriev, 1999), como se muestra en la figura 1.4.1

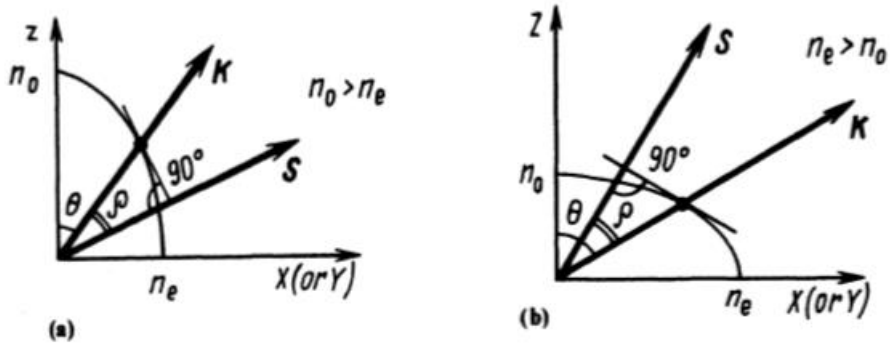

Figura 1.4.1. Ángulo de walk-off o de birrefringencia ρ para un material uniaxial, donde el vector $\vec{s}$ corresponde al vector de poynting y el vector $\vec{k}$ corresponde al vector de onda (a). Material uniaxial negativo (b). Material uniaxial positivo. (V.G. Dmitriev, 1999) (Capitulo 2).

Existe una relación entre el ángulo del eje óptico con respecto a la onda transmitida, el ángulo θ en la figura 1.4.1, y el ángulo de walk-off dada por la siguiente ecuación:

$$\rho(\theta) = \pm \arctan\left[\left(\frac{n_o}{n_e}\right)^2 \tan(\theta)\right] \mp \theta \tag{1.4.1}$$

Donde los signos superiores corresponden a materiales uniaxiales negativos y los inferiores a uniaxiales positivos. Como se muestra en la figura 1.4.2 para un material uniaxial negativo.

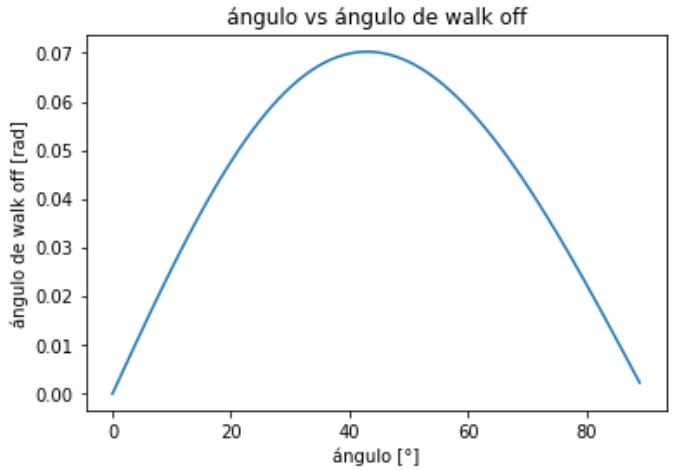

Figura 1.4.2. Gráfico de ángulo θ versus el ángulo de walk-off para un cristal de BBO a 1064 nm usando la ecuación (1.4.1), cuyos índices de refracción son $n_o = 1.6544$ y $n_e = 1.5424$ (G. Tamošauskas, 2018).

1.5 Propagación de una onda polarizada a través de un material uniaxial.

Consideremos la reflexión y transmisión de una onda en la interface entre dos medios uniaxiales como se muestra en la figura 1.5.1

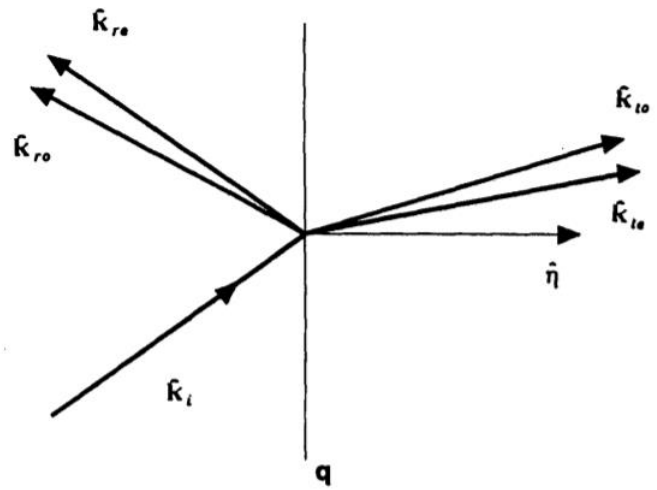

Figura 1.5.1. Geometría para refracción desde un material birrefringente hacia otro material birrefringente, Se generan dos rayos transmitidos (k_{to} y k_{te}) debido a la uniaxialidad del segundo materia, así como para los rayos reflejados y $\hat{\eta}$ es la normal a la interfaz, debido a que la primera interfaz es un material uniaxial, entonces también existen dos rayos reflejados al igual que en la refracción (Stephen C. McClain, (1992), Theory and physics, Section 3).

Tomando las relaciones constitutivas (1.1.6) y (1.1.7), para materiales uniaxiales la constante de permitividad eléctrica pasa a ser un tensor de la forma diagonal si es que el eje óptico del cristal se encuentra en el eje de propagación y tiene la siguiente forma

$$\epsilon = \begin{bmatrix} \epsilon_o & 0 & 0 \\ 0 & \epsilon_o & 0 \\ 0 & 0 & \epsilon_e \end{bmatrix} \tag{1.5.1}$$

Donde ϵ_o es la permitividad eléctrica dada por el índice de refracción ordinaria y ϵ_e es la permitividad eléctrica dada por el índice de refracción extraordinario. Si el material es quiral se usan las siguientes relaciones constitutivas (E.U. Condor , 1937):

$$\vec{D} = \epsilon\vec{E} + iG\vec{H} \tag{1.5.2}$$

$$\vec{B} = \mu\vec{H} - iG\vec{E} \tag{1.5.3}$$

donde G es el tensor gyrotropico del material y μ es la matriz identidad si es que el material es no magnético.

Tomando las ecuaciones constitutivas (1.5.2) y (1.5.3), junto con las ecuaciones de Maxwell (1.1.1) a (1.1.4), podemos resolver para $\vec{E}$ y obtener la siguiente ecuación de autovalores

$$M\vec{E} = 0 \tag{1.5.4}$$

donde definimos la matriz M como

$$M = \epsilon + (nK + iG)^2 \tag{1.5.5}$$

donde n corresponde al índice de refracción que está sintiendo la onda reflejada o transmitida y K es un operador producto cruz cuyos elementos se derivan del vector propagación $\vec{k}$ de la siguiente forma

$$K = \begin{bmatrix} 0 & -k_z & k_y \\ k_z & 0 & -k_x \\ -k_y & k_x & 0 \end{bmatrix} \tag{1.5.6}$$

si suponemos ahora que la forma del tensor gyrotropico es

$$G = \begin{bmatrix} g_0 & 0 & 0 \\ 0 & g_0 & 0 \\ 0 & 0 & g_e \end{bmatrix} \tag{1.5.7}$$

lo cual es válido para el cuarzo y también para materiales no ópticamente activos $(g_o = g_e = 0)$ siempre y cuando la orientación del cristal sea de tal forma que el eje óptico coincida con el eje de propagación del medio.

Si ahora la orientación del cristal es arbitraria los tensores ϵ y G se obtienen mediante sus tensores diagonales y una transformación unitaria de la siguiente forma

$$\epsilon = R^{-1}\epsilon_D R \tag{1.5.8}$$

$$G = R^{-1}G_D R \tag{1.5.9}$$

Donde ϵ_D y $G_D$son los tensores diagonales de las expresiones (1.5.1) y (1.5.7) respectivamente, y la transformación unitaria tiene la siguiente forma

$$R = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix} \begin{bmatrix} \cos\beta & \sin\beta & 0 \\ -\sin\beta & \cos\beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1.5.10}$$

Donde el eje óptico del cristal se encuentra a un ángulo α del eje $\hat{z}$, y la proyección del eje óptico sobre el plano $\hat{x}\hat{y}$ se encuentra a un ángulo β en el sentido anti horario del eje $\hat{x}$

Si escogemos como el plano de propagación de la onda electromagnética, el plano $\hat{y}\hat{z}$, entonces $\hat{k} = (0, \sin\theta, \cos\theta)$ y la ecuación de autovalores (1.5.5) queda como (Stephen C. McClain, (1992), Theory and physics)

$$\begin{vmatrix} \epsilon_o - g_o^2 - n^2 & -2ing_o\cos(\theta) & in(g_o + g_e)\sin(\theta) \\ 2ing_o\cos(\theta) & \epsilon_o - g_o^2 - n^2\cos^2(\theta) & -n^2\sin(\theta)\cos(\theta) \\ -in(g_o + g_e)\sin(\theta) & n^2\sin(\theta)\cos(\theta) & \epsilon_e - g_e^2 - n^2\sin^2(\theta) \end{vmatrix} E = 0 \quad (1.5.11)$$

donde el ángulo θ corresponde al ángulo entre el eje óptico y el vector de onda (figura (1.3.1)), y es dado por

$$\theta = arcos\left(\hat{k} \cdot \hat{c}\right) \quad (1.5.12)$$

Haciendo el determinante de la matriz anterior igual a cero obtenemos la siguiente ecuación cuadrática para n^2

$$n^4(e + w) - n^2[\epsilon_o'(\epsilon_e' + e) + v] + \epsilon_o'^2\epsilon_e' = 0 \quad (1.5.13)$$

donde

$$\epsilon_o' = \epsilon_0 - g_o^2 \quad (1.5.14)$$

$$\epsilon_e' = \epsilon_e - g_e^2 \quad (1.5.15)$$

$$e = \cos^2(\theta)\,\epsilon_e' + \sin^2(\theta)\,\epsilon_o' \quad (1.5.16)$$

$$v = 4\epsilon_e' g_o^2\cos^2(\theta) + \epsilon_o'(g_o + g_e)^2\sin^2(\theta) \quad (1.5.17)$$

$$w = (g_o + g_e)^2\sin^2(\theta)\cos^2(\theta) \quad (1.5.18)$$

Lo cual da las siguientes dos soluciones para n

$$n = \left\{\frac{\epsilon_o'(\epsilon_e' + e) + v \pm [\epsilon_o'^2(\epsilon_e' + e)^2 + 2\epsilon_o' v(\epsilon_e' + e) + v^2 - 4(e + w)\epsilon_o'^2\epsilon_e']^{\frac{1}{2}}}{2(e + w)}\right\}^{\frac{1}{2}} \quad (1.5.19)$$

para el signo superior en la ecuación (1.5.16) obtenemos el valor para n_e y para el signo inferior obtenemos el valor para n_o. Si ahora tomamos la ley de Snell de forma vectorial tenemos que (W.T.Welford, 1986):

$$\hat{\eta} \times \left(n_i\hat{k}_i\right) = \hat{\eta} \times \left(n_t\hat{k}_t\right) \quad (1.5.20)$$

donde $\hat{\eta}$ es la superficie normal que apunta hacia el medio transmitido y los sufijos i y t se refieren a incidente y transmitida respectivamente (figura (1.5.1)). La expresión 1.5.17 puede ser escrita de la siguiente forma (J.D. Trolinger, Jr, 1991):

$$n_t\hat{k}_t = n_i\hat{k}_i + \Gamma\,\hat{\eta} \quad (1.5.21)$$

donde Γ es llamada la constante astigmática, la cual se determina resolviendo la ecuación cuadrática que resulta de $\hat{k}_t \cdot \hat{k}_t$:

$$\Gamma = -n_i \hat{k}_i \cdot \hat{\eta} \pm \left[n_i^2 \left(\hat{k}_i \cdot \hat{\eta} \right)^2 + \left(n_t^2 - n_i^2 \right) \right]^{\frac{1}{2}} \tag{1.5.22}$$

usaremos el signo positivo para refracción y n_t es el índice de refracción del medio transmitido, mientras que el signo negativo se aplica para reflexión y n_t pasa a ser el índice de refracción del material por el cual se refleja la onda.

Usando la ecuación (1.5.19) en la ecuación (1.5.21) podemos obtener la orientación de la onda transmitida en un medio uniaxial, con un plano de incidencia que contenga al eje óptico.

El campo eléctrico total incidente como una función de tiempo y el espacio en el primer medio es:

$$\boldsymbol{E}_1 = \mathcal{R}e \left\{ \begin{matrix} E_i \{ \widehat{E}_\iota \exp\left[i \left(n_i \, \widehat{k_\iota} \cdot r - \omega t \right) \right] \} + a_{ro} \widehat{E}_{ro} \exp[i(n_{ro} \, \hat{k}_{ro} \cdot r - \omega t)] + \\ a_{re} \widehat{E}_{re} \exp[i(n_{re} \, \hat{k}_{re} \cdot r - \omega t)] \} \end{matrix} \right\} \tag{1.5.23}$$

donde $\boldsymbol{E}_1 = E_1 \widehat{E}_1$, a_{ro} es el coeficiente de Fresnel para el campo ordinario reflejado y a_{re} es el coeficiente de Fresnel para el campo extraordinario reflejado. El campo eléctrico total transmitido en el segundo medio es:

$$\boldsymbol{E}_2 = \mathcal{R}e \left\{ \begin{matrix} E_i \{ a_{to} \widehat{E}_{to} \exp[i(n_{to} \, \hat{k}_{to} \cdot r - \omega t)] + \\ a_{te} \widehat{E}_{te} \exp[i(n_{te} \, \hat{k}_{te} \cdot r - \omega t)] \} \end{matrix} \right\} \tag{1.5.24}$$

donde al igual que antes los subíndices to y te representan a los campo transmitidos ordinario y extraordinario respectivamente.

De forma similar podemos obtener la siguiente expresión para el campo magnético en la primera y la segunda interfaz

$$\boldsymbol{H}_1 = \mathcal{R}e \left\{ \begin{matrix} H_i \{ \widehat{H}_i \exp\left[i \left(n_i \, \widehat{k_\iota} \cdot r - \omega t \right) \right] \} + a_{ro} \widehat{H}_{ro} \exp[i(n_{ro} \, \hat{k}_{ro} \cdot r - \omega t)] + \\ a_{re} \widehat{H}_{re} \exp[i(n_{re} \, \hat{k}_{re} \cdot r - \omega t)] \} \end{matrix} \right\} \tag{1.5.25}$$

$$\boldsymbol{H}_2 = \mathcal{R}e \left\{ \begin{matrix} H_i \{ a_{to} \widehat{H}_{to} \exp[i(n_{to} \, \hat{k}_{to} \cdot r - \omega t)] + \\ a_{te} \widehat{H}_{te} \exp[i(n_{te} \, \hat{k}_{te} \cdot r - \omega t)] \} \end{matrix} \right\} \tag{1.5.26}$$

Desde las ecuaciones de Maxwell, los componentes de **E** y **H** tangenciales a la superficie son continuos a través de la superficie (Jackson John David, 1999). Si consideremos dos vectores ortogonales y tangenciales a la superficie, definidos como:

$$s_1 = \hat{k}_i \times \hat{\eta} \tag{1.5.27}$$

$$s_2 = \hat{\eta} \times s_1 \tag{1.5.28}$$

donde s_1 corresponde a la familia de vectores de polarizaciones –s, y s_2 a la familia de vectores de polarización –p. Aplicando condiciones de borde para las ecuaciones (1.5.23)-(1.5.26)

obtenemos un sistema lineal de cuatro ecuaciones relacionando los campos con los coeficientes de Fresnel de la forma

$$s_1 \cdot \left(\hat{E}_i + a_{ro}\hat{E}_{ro} + a_{re}\hat{E}_{re}\right) = s_1 \cdot \left(a_{to}\hat{E}_{to} + a_{te}\hat{E}_{te}\right) \tag{1.5.29}$$

$$s_2 \cdot \left(\hat{E}_i + a_{ro}\hat{E}_{ro} + a_{re}\hat{E}_{re}\right) = s_2 \cdot \left(a_{to}\hat{E}_{to} + a_{te}\hat{E}_{te}\right) \tag{1.5.230}$$

$$s_1 \cdot \left(\hat{H}_i + a_{ro}\hat{H}_{ro} + a_{re}\hat{H}_{re}\right) = s_1 \cdot \left(a_{to}\hat{H}_{to} + a_{te}\hat{H}_{te}\right) \tag{1.5.31}$$

$$s_2 \cdot \left(\hat{H}_i + a_{ro}\hat{H}_{ro} + a_{re}\hat{H}_{re}\right) = s_2 \cdot \left(a_{to}\hat{H}_{to} + a_{te}\hat{H}_{te}\right) \tag{1.5.32}$$

Los únicos valores desconocidos dentro de las ecuaciones anteriores son los coeficientes de Fresnel $a_{ro},\ a_{re}, a_{to}, a_{te}$. Estos valores pueden ser obtenidos solucionando las ecuaciones anteriores, obteniendo

$$\begin{bmatrix} a_{to} \\ a_{te} \\ a_{ro} \\ a_{re} \end{bmatrix} = \boldsymbol{F}^{-1} \cdot \begin{bmatrix} s_1 \cdot \hat{E}_i \\ s_2 \ \cdot \hat{E}_i \\ s_1 \ \cdot \hat{H}_i \\ s_2 \ \cdot \hat{H}_i \end{bmatrix} \tag{1.5.33}$$

donde

$$\boldsymbol{F} = \begin{bmatrix} s_1 \cdot \hat{E}_{to} & s_1 \cdot \hat{E}_{te} & -s_1 \cdot \hat{E}_{ro} & -s_1 \cdot \hat{E}_{re} \\ s_2 \cdot \hat{E}_{to} & s_2 \cdot \hat{E}_{te} & -s_2 \cdot \hat{E}_{ro} & -s_2 \cdot \hat{E}_{re} \\ s_1 \cdot \hat{H}_{to} & s_1 \cdot \hat{H}_{te} & -s_1 \cdot \hat{H}_{ro} & -s_1 \cdot \hat{H}_{re} \\ s_2 \cdot \hat{H}_{to} & s_2 \cdot \hat{H}_{te} & -s_2 \cdot \hat{H}_{ro} & -s_2 \cdot \hat{H}_{re} \end{bmatrix} \tag{1.5.34}$$

Con la ecuación (1.5.33) obtenemos los coeficientes de Fresnel para los campos transmitidos y reflejados, usando estos campos en las ecuaciones (1.5.23)-(1.5.26) obtenemos los campos eléctricos y magnéticos para la primera y segunda interfaz del problema.

1.6 Algoritmo para obtener los coeficientes de Fresnel usando dos interfaces.

Supongamos que ahora el problema se centra en tres interfaces, donde la primera y la tercera son interfaces isotrópicas y la segunda es birrefringente. Como se muestra en la siguiente figura

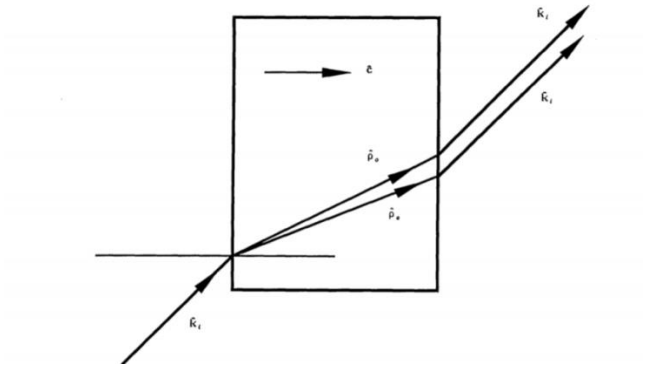

Figura 1.6.1. Problema de un rayo refractando desde un material isotrópico hacia un material uniaxial, que luego vuelve al material isotrópico anterior. Donde $\hat{c}$ es el eje óptico del sistema. (Stephen C. McClain, (1992), Algorithms, Section 3).

Lo primero que debemos hacer para abarcar este problema es separarlo en dos, primero obtendremos las refracciones y la reflexión del rayo pasando desde la primera interfaz (isotrópica) hacia la segunda interfaz (birrefringente), y luego con estos resultados, analizaremos el caso del rayo pasando desde la segunda interfaz hacia la tercera interfaz, ósea el caso birrefringente-isotrópica.

1.6.1 Caso isotrópico-Birrefringente

En la sección 1.5 introducimos las ecuaciones que usaremos para obtener los coeficientes de Fresnel, pero las ecuaciones (1.5.16) y (1.5.19) se usan mutuamente debido a la ecuación usada para obtener el ángulo θ (1.5.9), por lo que usaremos un algoritmo de convergencia para las ecuaciones (1.5.16) y (1.5.9) de la siguiente forma (Stephen C. McClain, (1992), Algorithms):

1) Escogemos $n = \sqrt{\epsilon_o}$ como estimado inicial, usando la constante dieléctrica ordinaria del medio en el cual es rayo es propagado.

2) Usamos la ecuación (1.5.19) tanto para el rayo transmitido como para el reflejado.

3) Calculamos $\hat{k}$ transmitido de la ecuación (1.5.18)

4) Obtenemos el ángulo θ de la ecuación (1.5.9)

5) Calculamos un nuevo valor para n usando la ecuación (1.5.16) donde los valores de la constante dieléctrica y las constantes giro trópicas son calculadas del medio en el cual es rayo se propaga.

6) Repetimos los pasos 2-5 hasta que n converga.

7) Usamos los últimos valores obtenidos en el paso 3 para $\hat{k}$ y en el paso 5 para el n tanto para los rayos reflejados como transmitidos.

8) Calculamos con la ecuación $(1.5.19)$, usando el signo superior y los valores obtenidos para n_o, el índice de refracción extraordinario y luego usando la ecuación (1.5.18) obtenemos el $\hat{k}$ para el rayo extraordinario.

9) Usamos la ecuación $(1.5.5)$ para resolver la ecuación $(1.5.4)$ para así obtener los campos transmitidos.

10) Para los campos reflejados primero debemos obtener $\hat{k}_r$ usando la siguiente ecuación

$$\hat{k}_r = \frac{n_i\hat{k}_i - 2n_i(\hat{k}_i \cdot \hat{\eta})\hat{\eta}}{\left|n_i\hat{k}_i - 2n_i(\hat{k}_i \cdot \hat{\eta})\hat{\eta}\right|} \tag{1.6.1.1}$$

11) Usando el valor obtenido para $\hat{k}_r$ podemos obtener los campos reflejados usando:

$$\hat{E}_{rs} = \frac{\hat{k}_r \times \hat{\eta}}{\left|\hat{k}_r \times \hat{\eta}\right|} \tag{1.6.1.2}$$

$$\hat{E}_{rp} = \hat{k}_r \times \hat{E}_{rs} \tag{1.6.1.3}$$

$$\hat{H}_{rs} = n_i\hat{k}_r \times \hat{E}_{rs} \tag{1.6.1.4}$$

$$\hat{H}_{rp} = n_i\hat{k}_r \times \hat{E}_{rp} \tag{1.6.1.5}$$

12) Calculamos el vector de campo magnético usando:

$$\hat{H}_{uv} = (n_{uv}K_{uv} + iG)\hat{E}_{uv} \tag{1.6.1.6}$$

Donde los sufijos uv pueden ser reemplazados por $to, te\ , ro, re$

13) Calculamos el vector de campo magnético incidente usando:

$$H_i = (n_iK_i + iG_i)\hat{E}_i \tag{1.6.1.7}$$

14) Determinamos el vector de Pointing usando la siguiente relación(Jackson John David, 1999):

$$\hat{\rho}_{uv} = \frac{\mathcal{R}e\{\hat{E}_{uv} \times \hat{H}_{uv}{}^{*}\}}{|\mathcal{R}e\{\hat{E}_{uv} \times \hat{H}_{uv}{}^{*}\}|} \tag{1.6.1.8}$$

15) Al obtener todos los campos usamos la ecuación (1.5.30) solo que ahora no para dos interfaces uniaxiales, sino que para la situación isotrópico-birrefringente, quedando de la siguiente forma:

$$\begin{bmatrix} a_{to} \\ a_{te} \\ a_{rs} \\ a_{rp} \end{bmatrix} = \boldsymbol{F}^{-1} \cdot \begin{bmatrix} s_1 \cdot \hat{E}_i \\ s_2 \cdot \hat{E}_i \\ s_1 \cdot \hat{H}_i \\ s_2 \cdot \hat{H}_i \end{bmatrix} \tag{1.6.1.9}$$

Dónde:

$$\boldsymbol{F} = \begin{bmatrix} s_1 \cdot \hat{E}_{to} & s_1 \cdot \hat{E}_{te} & -s_1 \cdot \hat{E}_{rs} & 0 \\ s_2 \cdot \hat{E}_{to} & s_2 \cdot \hat{E}_{te} & 0 & -s_2 \cdot \hat{E}_{rp} \\ s_1 \cdot \hat{H}_{to} & s_1 \cdot \hat{H}_{te} & 0 & -s_1 \cdot \hat{H}_{rp} \\ s_2 \cdot \hat{H}_{to} & s_2 \cdot \hat{H}_{te} & -s_2 \cdot \hat{H}_{rs} & 0 \end{bmatrix} \tag{1.6.1.10}$$

16) Calculamos el optical path lenght (OPL) del rayo que pasa por la segunda interfaz, calculando primero la longitud física o physical lenght l:

$$l = [(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{\frac{1}{2}} \tag{1.6.1.11}$$

Usando este valor podemos calcular el OPL de la siguiente forma:

$$OPL = nl\hat{k} \cdot \hat{\rho} \tag{1.6.1.12}$$

1.6.2 Caso isotrópico-Birrefringente-isotrópico.

Para este caso usaremos los valores obtenidos en la sección 1.6.1, y analizaremos el caso del rayo transmitido desde una interfaz birrefringente hacia una isotrópica.

El caso es similar al planteado en la sección 1.6.1 solo que ahora tenemos dos rayos reflejados por lo que se obtendrá $\hat{k}_{ro}$ y $\hat{k}_{re}$, y solo un rayo transmitido $\hat{k}_t$, por lo que se seguirá el mismo algoritmo solo que con algunos cambios mencionados a continuación (Stephen C. McClain, (1992), Algorithms) :

1) Para calcular $\hat{k}_t$ usaremos la siguiente ecuación:

$$\Gamma_t = -n_i\hat{k}_i \cdot \hat{\eta} \pm \left[n_i^2\left(\hat{k}_t \cdot \hat{\eta}\right)^2 + \left(n_t^2 - n_i^2\right)\right]^{\frac{1}{2}} \tag{1.6.2.1}$$

$$\hat{k}_t = \frac{n_i\hat{k}_i + \Gamma_t\,\hat{\eta}}{\left|n_i\hat{k}_i + \Gamma_t\,\hat{\eta}\right|} \tag{1.6.2.2}$$

2) Los campos transmitidos serán obtenido usando:

$$\hat{E}_{ts} = \frac{\hat{k}_t \times \hat{\eta}}{\left|\hat{k}_t \times \hat{\eta}\right|} \tag{1.6.2.3}$$

$$\hat{E}_{tp} = \hat{k}_t \times \hat{E}_{ts} \tag{1.6.2.4}$$

$$\hat{H}_{ts} = n_i\hat{k}_t \times \hat{E}_{ts} \tag{1.6.2.5}$$

$$\hat{H}_{tp} = n_i\hat{k}_t \times \hat{E}_{tp} \tag{1.6.2.6}$$

3) Por ultimo obtendremos los coeficientes de Fresnel de la siguiente ecuación:

$$\begin{bmatrix} a_{ts} \\ a_{tp} \\ a_{ro} \\ a_{re} \end{bmatrix} = \boldsymbol{F}^{-1} \cdot \begin{bmatrix} s_1 \cdot \hat{E}_i \\ s_2 \cdot \hat{E}_i \\ s_1 \cdot \hat{H}_i \\ s_2 \cdot \hat{H}_i \end{bmatrix} \tag{1.6.2.7}$$

donde

$$\boldsymbol{F} = \begin{bmatrix} s_1 \cdot \hat{E}_{ts} & 0 & -s_1 \cdot \hat{E}_{ro} & -s_1 \cdot \hat{E}_{re} \\ 0 & s_2 \cdot \hat{E}_{tp} & -s_2 \cdot \hat{E}_{ro} & -s_2 \cdot \hat{E}_{re} \\ 0 & s_1 \cdot \hat{H}_{tp} & -s_1 \cdot \hat{H}_{ro} & -s_1 \cdot \hat{H}_{re} \\ s_2 \cdot \hat{H}_{ts} & 0 & -s_2 \cdot \hat{H}_{ro} & -s_2 \cdot \hat{H}_{re} \end{bmatrix} \tag{1.6.2.8}$$

Después de obtener los coeficientes de Fresnel de ambas interfaces se calculara el campo total transmitido después de que el rayo pasara por ambas interfaces, primero calcularemos el optical path diference (OPD) entre los rayos ordinarios y extraordinarios dentro del material uniaxial usando

$$OPD = OPL_e - OPL_o + (y_o - y_e)\sin(\theta_i) \tag{1.6.2.9}$$

Donde los OPL son obtenidos por la ecuación (1.6.1.12) tanto para el rayo extraordinario como el ordinario y $y_o - y_e$ corresponde al corrimiento horizontal (usado como $y_o - y_e$, debido a que el problema se planteó inicialmente en el plano y-z). Ahora con este valor podemos obtener el desfase espacial que se tiene un rayo con otro (ordinario con el extraordinario) usado:

$$\Delta\phi = \frac{OPD}{\lambda_o} 2\pi \tag{1.6.2.10}$$

Finalmente podemos calcular el valor del campo total transmitido en la tercera interfaz, que vendría siendo un overlap entre los dos rayos transmitidos, usando (Stephen C. McClain, (1992), Algorithms) :

$$\hat{E}_{out} = \begin{Bmatrix} [\alpha_{1tos} \cdot \alpha_{2tos} + \alpha_{1tes} \cdot \alpha_{2tse} \exp(-i\Delta\phi)]\hat{s} \\ +\left[\alpha_{1tos} \cdot \alpha_{2tpo} + \alpha_{1tep} \cdot \alpha_{2tpe} \exp(-i\Delta\phi)\right]\hat{p} \end{Bmatrix} \times \exp\left[i\left(\hat{k}_t \cdot r - \omega t\right)\right] \tag{1.6.2.11}$$

Capítulo 2: Regresión de datos con aprendizaje de máquina supervisado

2.1 Aprendizaje de máquina (Machine Learning).

Machine Learning generalmente se refiere a cualquier cambio en un sistema que ejecuta una tarea que tenga alguna asociación con una inteligencia artificial (AI por sus siglas en inglés), dichas tareas pueden involucrar reconocimiento, diagnósticos, planificación, control robótico, predicciones, entre otras cosas (Nilsson N.J. 1996). Para que un sistema sea inteligente, este debe ser capaz de adaptarse a en su entorno. Esta habilidad es llamada aprendizaje (Ethem Alpaydin, 2004). Dentro de los algoritmos de Machine Learning existen tres grandes categorías Unsupervised Learning, Reinforcement Learning y Supervised Learning.

Un algoritmo de Supervised Learning trata de encontrar un mapeo entre un input y un output que están etiquetados de esta forma por un "supervisor", de ahí su nombre. Un algoritmo de Unsupervised Learning solo necesita inputs, por lo que no existe un "supervisor" que lo asocie con outputs. El objetivo de este algoritmo es encontrar elementos de similitud o regularidade en los datos de input. Por último, un algoritmo de Reinforcement Learning se basa en un tipo de output específicos, los cuales deben ser sistemas de acciones. El algoritmo se centra primordialmente en cuál es la mejor secuencia de acciones que se deben llevar a cabo para lograr un objetivo definido. Esta secuencia es llamada *policy* y la mejor forma de entender este tipo de algoritmo es considerar un juego de video donde quizás una acción determinada no implica ningún cambio en el juego pero sí una gran cantidad de acciones pueden llevar a un puntaje máximo (Ethem Alpaydin, 2004).

Existen varios algoritmos básicos que se utilizan en la implementación de esquemas de Machine Learning e inteligencia artificial, como por ejemplo Decision Tree Learning, Genetic Algorithms, Bayesian Learning y Aritificial Neural Networks (ANN). En esta tesis utilizaremos Artificial Neural Networks (Tom M. Mitchell, 1997).

2.2 Artificial Neural Networks (ANN).

El estudio de las ANNs ha sido inspirado en parte por la observación biológica del aprendizaje humano que se tiene en las redes neuronales, debido a esto las neuronas de las ANNs, globalmente hablando, funcionan parecido a cómo funciona una neurona biológica: Ante un estímulo que supera un umbral la neurona se activa. Más tarde hablaremos la implicancia que tiene que la neurona se active y cuál es este umbral que se debe sobrepasar para ser estimulada.

La forma más simple de una ANN es un Perceptron, ilustrado en la figura 2.2.1. El Perceptron toma un vector de valores reales $\{x_1, x_2, \ldots, x_n\}$, calcula una combinación lineal de estos vectores usando un conjunto de pesos $\{W_1, W_2, \ldots, W_n\}$ y con esto entrega un output. El output en este caso es 1 si se cumple una cierta condición dada por valores predefinidos y -1 si es que no lo cumple.

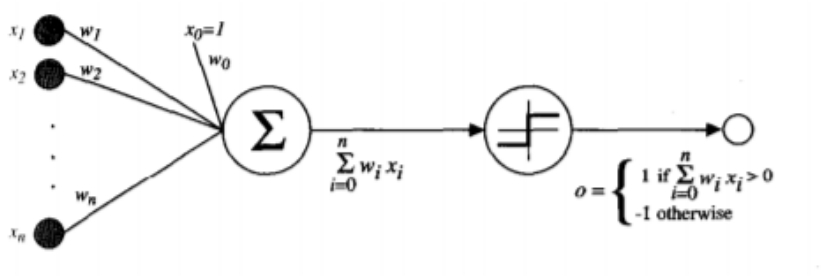

Figura 2.2.1. Ilustración de un Perceptron, desde izquierda a derecha se encuentran los inputs, la combinación lineal de los vectores de input, la función delimitante y por último el output (Tom M. Mitchell, 1997) (Capitulo 4).

El Perceptron consta de cuatro partes bien definidas las cuales pueden ser alteradas para un mejor manejo de lo que se quiera hacer con este, la primera son los valores de inputs, que como se mencionó anteriormente deben ser valores reales, luego están la combinación lineal de los vectores de input donde a cada valor de input se le multiplica por peso, luego esta suma pasa por una función que en el caso anterior fue una función señal. Si el valor de es mayor que cero entonces entrega como salida un 1 y en otro caso entrega como salida un -1. Esta decisión se implementa con una función delimitante. Por ultimo está el output, que es el valor que uno obtiene después de ejecutar todo el ciclo antes mencionado. Mediante este esquema, un Perceptron fácilmente puede representar funciones booleanas como OR, AND y XOR.

Para entrenar un Perceptron, debemos tener varios datos de entrada y salida. Entre más datos se tiene, mejor va a ser el entrenamiento. Los datos de entrenamiento son vectores de entrada asociados asu respectivo valor de output. Durante el entrenamiento del Perceptron, se van alterando los valores iniciales de los pesos para cada dato de la siguiente forma.

$$w_i = w_i + \Delta w_i \,, \tag{2.2.1}$$

Donde el lado izquierdo corresponde al nuevo peso obtenido para todos los valores de input, después de una entrada de datos, en la parte derecha están los pesos anteriores y un ajuste agregado (Δw_i) de forma que los valores de los pesos sean más representativos para el dato en cuestión (Tom M. Mitchell, 1997).

Un artificial neural network ANN, se basa en el funcionamiento del Perceptron pero aumenta la cantidad de "neuronas" o la cantidad de perceptrones que tiene el sistema. En el caso de las ANN también tenemos un vector de entrada solo que ahora en vez de tener una neurona o un solo bloque de procesamiento, como los perceptrones. Ahora puede haber N bloques, donde cada bloque se activa usando una función de activación definida por el usuario. Estas funciones sirven como delimitantes, debido a que si superan un umbral numérico dado por los pesos de cada neurona y los valores de input, entonces las neuronas en el bloque se activan simultáneamente, haciendo que sus pesos tengan influencia sobre los outputs. Si las neuronas no se activan, entonces no tendrán influencia sobre el output para los valores dados en input. Si bien tenemos más de una neurona en la red, también podemos generar varias filas de neuronas las cuales estarán interconectadas entre ellas como se muestra en la figura (2.2.2). Estas filas de neuronas son conocidas como hidden layers o capas ocultas, debido a que el usuario sólo tiene control de los valores iniciales de estas capas, como los pesos de estas pero luego debido al entrenamiento estos valores varían internamente y se pierde el control externo de estas variables. Luego de este proceso de ajuste de pesos. se generan los output dados por la aplicación sobre los inputs de cada una de las neuronas activadas. Al igual que en el caso de los perceptrones, los valores que se entrenan dentro de una ANN son los pesos y además se agregara un valor más que son los *Bias.* Estos corresponden al valor que cada neurona agrega al dato de input para transformarlo en el output (Tom M. Mitchell, 1997).

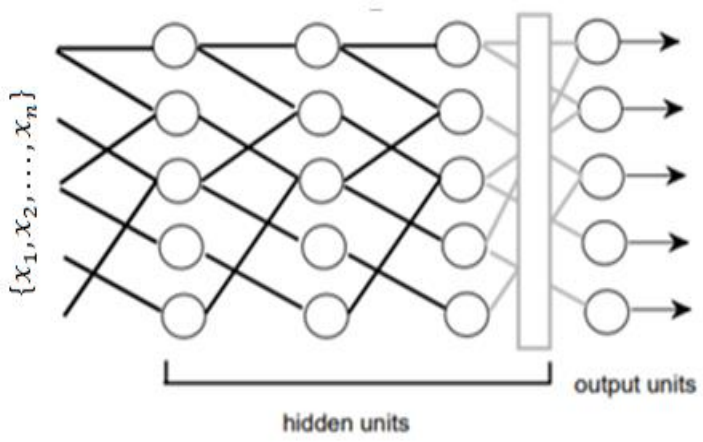

Figura 2.2.2. Ilustración de una ANN. Desde izquierda a derecha se encuentran los inputs, los hidden layers o hidden units, y por último el output (Tom M. Mitchell, 1997) (Capitulo 4).

Cabe mencionar que un Perceptron, debido a su simplicidad, está diseñado para entender o representar funciones lineales, mientras que una artificial neuronal network puede representar fácilmente una función no lineal y la complejidad de la ANN debe ser proporcional a la complejidad del problema que se está estudiando. Esto quiere decir que si la función que se quiere entender o representar es compleja entonces la red neuronal deberá tener una mayor cantidad de hidden layers y una mayor cantidad de neuronas que una red neuronal que busque representar una función menos compleja.

2.3 Feed-Forward Neural Networks.

Una de las redes neuronales más simples son las llamadas Feed-Forward Neural Networks. Estas tienen un diseño que funciona de izquierda a derecha, como se muestra en la figura 2.2.1 y su nombre hace referencia a esto. Para evaluar el rendimiento que tiene una ANN en términos del aprendizaje deseado, existen varias variables que pueden usarse para validar el entrenamiento de la red. Por ejemplo, se puede validar si es que la red neuronal es capaz de comprender y representar la función no lineal que se busca dentro del problema.

El método de entrenamiento para este tipo de redes es similar al de los perceptrones mencionados con anterioridad, solo que en este caso son dos las variables por neurona que deben ir adecuándose al problema: el peso de cada input y los *Bias* de cada neurona, es decir si entra un vector de datos dentro de una feed-forward neural network (figura 2.2.1), después se actualizan los pesos y *bias* de cada neurona de la red de derecha a izquierda, y por último se obtiene un output que se compara con el valor del output provisto por el supervisor, en un esquema de Supervised Learning.

Dentro de una red neuronal existen varias variables que se pueden ajustar para facilitar tanto el aprendizaje de la máquina como para optimizar los tiempos de cálculos de esta. Entre estas variables están algunas ya mencionadas como el número de neuronas y el número de hidden layers. Otras variables importantes son las siguientes:

1) **Data sets**: nos referiremos a un data set como todos los datos de entrada y salida usados para entrenar una red neuronal.

2) **Epochs (épocas)**: son la cantidad de veces que toda la data set es usada para entrenar la red neuronal.

3) **Learning rate**: es la tasa a la cual la red neuronal va variando sus valores para adecuarse a la data que se le está entregando en ese momento. Es decir, es una taza de aprendizaje. Entre más grande es este valor, más varían las variables de la red y viceversa.

4) **Batch size**: es la cantidad de datos que se usan dentro del data set para ir actualizando las variables de la red. Es decir, es la porción de datos que se usa para ir actualizando la red neuronal, normalmente el *default* es de uno pero cuando se trabajan con mayor cantidad de datos se debe aumentar para disminuir el tiempo de cálculo.

5) **Función de activación**: esta función es la misma que llamamos antes como función delimitante, la cual tiene la función de activar la neurona si es que un el valor de los pesos por los inputs traspasa un cierto límite. Algunas de las funciones más usadas son la función signo, la función lineal, la exponencial lineal y la sigmoidea.

6) **Predicción del modelo**: corresponde a la predicción hecha por una red neuronal, la cual fue entrenada previamente, obtenida para valores de inputs dados.

7) **Función de pérdida**: esta función busca comparar las predicciones obtenidas para los datos usados para entrenar la red y los datos de salida de estos, esta función es también llamada función de error ya que calcula un error relativo entre los valores introducidos en la red y los valores calculados por esta.

8) **Optimizador**: corresponde a una función que busca optimizar el aprendizaje de la red neuronal usando los datos de la función de pérdida, ósea los datos predichos y los datos introducidos en la red neuronal, esta función determina la convergencia de la red neuronal.

Si éstas variables no son elegidas correctamente, dentro del proceso de aprendizaje pueden ocurrir dos fenómenos estos son el underfitting y el overfitting.

Underfitting se refiere a cuando una red neuronal no logra adecuarse a la función que busca el problema por lo que, o se deberá hacer una red un poco más "profunda" (lo que implica mayor cantidad de neuronas y de *hidden layers*), o cambiar algunos parámetros como el *learning rate* o entrenar por más tiempo la red. Si la profundidad de la red no es suficiente para capturar bien la nolinealidad del problema, ninguna variable puede garantizar que no exista underfitting.

Overfitting se refiere a cuando una red neuronal sobreestima los valores. Esto quiere decir que la red aprende más de lo que debería sobre el problema y genera relaciones entre input y output que no vienen dadas por la nolinealidad del problema. Existen dos casos uno en que la red que se aprenda de "memoria" los valores de input y output del problema, y otro en que encuentre una relación menos complicada del problema entre la data set. Contrario al underfitting, ésto puede ocurrir porque la red es muy profunda en comparación con el problema que se desea resolver o que las variables no estén bien definidas para el problema, como por ejemplo que los datos de la data set no estén bien distribuidos dentro del problema. Esto se puede evitar probando la red para una muestra más pequeña de datos y tomando datos de forma aleatoria para conformar la data set.

Para saber si es que estamos en presencia de alguno de estos dos problemas, especialmente el problema de overfitting que es el más complejo de detectar, normalmente se usa una herramienta llamada *validation data* o datos de validación. La *validation data* corresponde a tomar el data set y cortarlo de forma que un porcentaje alto corresponda a un nuevo data set, y que el otro porcentaje restante corresponda a una validation data (generalmente entre un 80-20 % de los datos totales). Luego de que se entrene la red con la data set y en vez de preocuparse por la función de pérdida de la data set, se buscan los valores de la función de pérdida de la validation data, ya que como se mencionó anteriormente estos datos no fueron usados para entrenar la red y así obtener una estimación más precisa del error que se va a obtener para una futura predicción.

Otra opción para evitar el overfitting es el uso de otra herramienta llamada *Dropout*, esta herramienta hace que no todas las neuronas de un hidden layer se activen cuando se va a hacer el entrenamiento de estas neuronas. Esto funciona asignándole un porcentaje de dropout o de abandono a cada hidden layer, normalmente este porcentaje oscila entre un 20% y un 50%, donde si es por debajo del 20% no tendrá ningún efecto para redes neuronales complejas y sobre un 50% tenderá a generar underfitting al momento de usar la validation data. Las neuronas que no van a ser tomadas en cuenta en cada instancia de entrenamiento se eligen de manera aleatoria, por lo que que si alguna neurona o paquetes de neuronas dentro de una hidden layer esté generando overfitting, simplemente parte de este paquete o la neurona en sí misma no siempre será tomada en cuenta para el entrenamiento.

2.4 Machine Learning usando Python.

Para poder implementar un algoritmo de Machine Learning usaremos Python como lenguaje de programación, tomando como compilador principal Spyder de Anaconda. Para esta implementación necesitaremos dos librerías principales de redes neuronales las cuales son **Tensorflow** y **Keras**. A continuación se dará una lista de los pasos que siguieron en esta tesis para instalar las librerías usadas en GPU.

1) Instalar anaconda siguiendo los pasos dados en https://docs.anaconda.com/anaconda/install/.

2) Instalar Cuda Toolkit.

3) Descar e instalar cuDNN de la página de NVIDIA.

4) Crear un nuevo ambiente.

5) Abrir una cmd o pantalla de comando donde activamos el ambiente que creamos en el paso 4.

6) Instalar en ese nuevo ambiente la libreria Tensorflow usando “conda install tensorflow-gpu”.

7) Instar Keras usando “pip install keras”.

Se instalan las dos librerías debido a que Keras corresponde a una sub-libreria de Tensorflow, por lo que si Tensorflow no está instalado, Keras no puede ser usado. Pero para esta tesis usaremos Keras para hacer cualquier algoritmo que tenga alguna incumbencia en las redes neuronales.

Para crear una red neuronal en Keras debemos usar las siguientes funciones de la librería (Choller, 2005):

1) **Input**: corresponde al inicio de una red neuronal y a la corrida de inputs que se quiere tener, por ejemplo Input (shape=(16,)) significa que tendremos 16 valores para el vector de input de la red.

2) **BatchNormalization**: es una herramienta que permite normalizar los valores de entrada de la red, esto se usa debido a que las redes neuronales trabajan de mejor forma con valores que oscilan entre 0 y 1, y esta herramienta permite que la red neuronal converja más fácilmente. Esta función se debe usar después de definir cada hidden layer de la red.

3) **Dense**: corresponde a un Hidden Layer y se usa especificando la cantidad de neuronas que tendrá cada Hidden layer, por ejemplo Dense (300) corresponde a una hidden layer con 300 neuronas, estas neuronas por defecto están conectadas totalmente a la Hidden layers que las sucede.

4) **Concatenate**: corresponde a una concatenación de varias hidden layers, normalmente se usa para generar mayor profundidad en la red sin aumentar la cantidad de neuronas que esta tiene.

5) **Dropout**: corresponde al porcentaje de abandono que tendrá la hidden layer al hacer el entrenamiento de la red, se usa de la siguiente manera Dropout(0.3) donde en este caso especifica un Dropout de un 30%.

6) **Activation**: especifica la función de activación de cada neurona dentro de una hidden layer. Donde las opciones posibles son las siguientes: softmax, elu, selu, softplus, softsign, relu, tanh, sigmoid, hard sigmoid, exponential y linear.

7) **Model**: es una herramienta para crear una red neuronal, la cual nos permite definir el inicio y el final de la red que deseamos crear.

8) **Compile**: es una sub rutina de Model la cual nos permite compilar la red neuronal con las especificaciones que nosotros queramos.

9) **Losses**: dentro de la subrutina de Compile, con esta función podemos especificar el tipo de función de pérdida que deseamos que se calcule dentro de la red.

10) **Optimizer**: dentro de la subrutina de Compile, con esta función podemos especificar el tipo de optimizador que deseamos que se utilice dentro de la red neuronal.

11) **Metrics**: dentro de la subrutina de Compile, permite especificar lo que se quiera optimizar dentro de la red, generalmente corresponde a funciones como mínimos cuadrados, error relativo porcentual, accuracy, pérdida, etc.

12) **Fit**: es una subrutina de Model la cual nos permite entrenar la red neuronal ya creada con las funciones antes mencionadas, en esta subrutina debemos colocar como input el data set, es decir, los vectores de input y los vectores de output de la red y además estos mismos vectores pero de validación.

13) **Epochs**: es una subrutina de fit, la cual nos permite decidir cuantas veces repetimos el entrenamiento entero dentro de una red usando todo el data set.

14) **Batch_size**: es una subrutina de fit, la cual nos permite elegir con cuantos datos se actualizara los valores de la red neuronal, por ejemplo podemos tener 5000 datos dentro de la data set, pero queremos actualizar la red neuronal cada 300 datos por lo tanto usamos batch_size=300.

15) **Predict**: es una subrutina de model, la cual está encargada de predecir valores después de output, usando un valor de input definido, luego de que la red sea entrenada, este proceso normalmente se usa para saber si es que la red está bien entrenada.

Usando todas las funciones mencionadas con anterioridad podemos crear nuestra propia red neuronal, con cual tenemos varias opciones que podemos modificar como solo generar hidden layers densas, colocar algunas concatenadas, cambiar el función de activación, cambiar el compilador, cambiar la función de optimización, entre otras modificaciones.

2.5 Caso ejemplo: interpolación nolineal con una red neuronal.

Como se mencionó con anterioridad, la idea de trabajar con redes neuronales es su potencial para hacer regresiones o interpolaciones sobre funciones altamente nolineales, por lo que a modo de ejemplo tomaremos una función trigonométrica y haremos un análisis para detectar que tan buena es la regresión nolineal hecha por una red neuronal simple. Esto nos permitirá entender el funcionamiento de una red neuronal, y su potencial de escalamiento a funciones más complejas.

Para este caso se utilizó la función trigonométrica coseno, con 100.000 datos de entrenamiento y 30.000 datos de validación, con una red neuronal que tiene las siguientes propiedades: son 6 hidden layers, las cuales están conectadas totalmente entre ellas, donde la primera tiene 100 neuronas, la segunda tiene 75 neuronas, la tercera 50 neuronas, la cuarta 25 neuronas, la quinta 10 neuronas y la última es la salida que tiene una neurona. Donde todas las hidden layer (excepto la de salida) tiene un 30% de Dropout y una función de activación "elu". Con una función de optimización "Adam" para la red, con un Learning rate de 0.1 y un decaimiento de este por época de un 0.01, una función de pérdida del tipo logcosh y con una métrica a medir de mean square error o error cuadrático medio, en la época 30 se cambió el optimizador por "SGD" (que converge de mejor forma cuando el error es menor, que el "Adam"). La red fue entrenada en 87 epocas, con un tiempo de computación promedio de unos 21 $[s]$ por set de entrenamiento. Cabe mencionar que la red neuronal solo se alimentó con datos de ángulos entre 0 y π [rad] (o de 0 a 180 grados).

Con la red neuronal antes mencionada se obtuvieron los siguientes resultados para la interpolación no lineal de la función trigonométrica coseno:

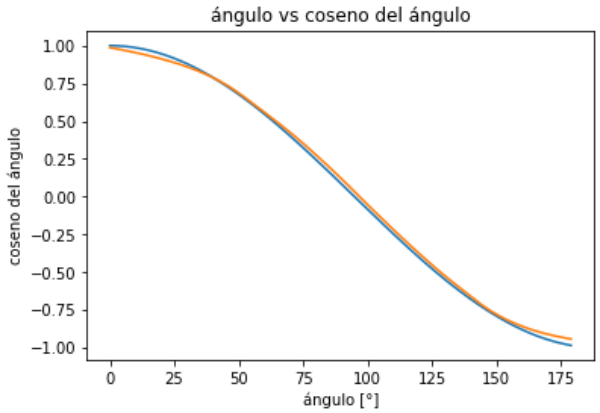

Figura 2.5.1. Gráfico de ángulo vs coseno del ángulo, para la red neuronal mencionada en este capítulo, este grafico va de 0 a π radianes.

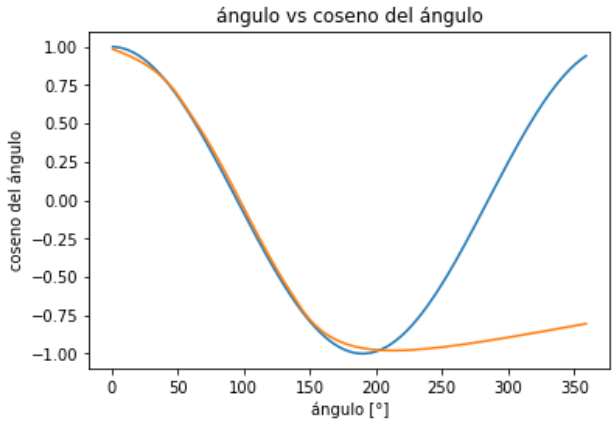

Figura 2.5.2. Gráfico de ángulo vs coseno del ángulo, para la red neuronal mencionada en este capítulo, este grafico va de 0 a 2π radianes.

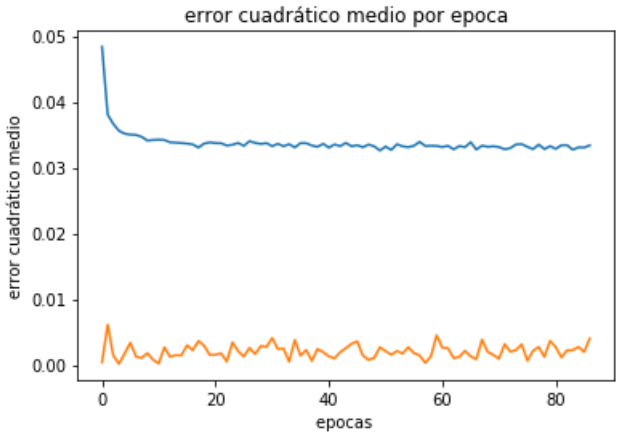

Figura 2.5.3. Gráfico del error cuadrático medio por época para la red neuronal mencionada en este capítulo, desde el segundo paso hasta el paso 150 (el ultimo).

Como se puede ver del gráfico de la figura 2.5.1 podemos ver que la regresión lineal fitea de buena forma la función coseno, con algunos problemas en los valores cercanos a cero, esto se puede deber a la que como la generación de datos fue aleatoria la cantidad de datos para entrenar en este intervalo pudo haber sido menor que en algún otro intervalo, pero en general la interpolación para este problema es consistente con la función coseno. Si observamos el gráfico de la figura 2.5.2 podemos ver que la red neuronal solo hace una interpolación consistente entre los valores que fue entrenada la red, en este caso desde los 0° hasta los 90°, después de este valor la red neuronal no es consistente por lo que sus predicciones no deben tomarse en cuenta. Del grafico de la figura 2.5.3 se puede observar que el error cuadrático medio baja bruscamente después de la 6 época y luego llega a una meseta en la cual la interpolación no varía mucho, teniendo ciertos picks crecientes en algunas épocas sobre todo para los datos de validación, este comportamiento es típico para cualquier red neuronal.

Se obtuvieron los siguientes valores para la red neuronal usada en la interpolación del coseno con un batch size de 100 datos por actualización de red: la pérdida para los datos de entrenamiento es de 0.0164, el error cuadrático medio para estos mismos es de 0.0334, para los datos de validación tenemos que la pérdida es de 0.0020 y el error cuadrático medio es de 0.0040.

Capítulo 3: Elipsometría de Mueller usando redes neuronales pre-alimentadas (FFNN)

3.1 Elipsometría de Mueller

La elipsometría de Mueller se centra en vectores, llamados vectores de Stokes, que caracterizan la polarización de una onda y una matriz de transición, llamada matriz de Mueller.

Supongamos que al igual que en el caso de la elipsometría de Jones tenemos una onda plana monocromática viajando en la dirección z. Entonces podemos separar la onda en los componentes E_x y E_y de la siguiente forma

$$E_x = a_x e^{i\varphi_x} \tag{3.1.1}$$

$$E_y = a_y e^{i\varphi_y} \tag{3.1.2}$$

Por lo tanto definiremos el vector de Stokes de la siguiente forma:

$$\vec{S} = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \tag{3.1.3}$$

donde

$$\begin{aligned} S_0 &= \quad |E_x|^2 + |E_y|^2 \\ S_1 &= \quad E_x E_y^* + E_x^* E_y \\ S_2 &= i(E_x E_y^* - E_x^* E_y) \\ S_3 &= \quad |E_x|^2 - |E_y|^2 \end{aligned} \tag{3.1.4}$$

Ahora supongamos que tenemos el caso en que una onda pasa a través de un material, entonces los vectores de Stokes de entrada y salida están relacionados por una matriz de transición llamada matriz de Mueller:

$$\vec{S}_{out} = M\vec{S}_{in} \tag{3.1.5}$$

Esta matriz se puede obtener experimentalmente usando los métodos de elipsometría mencionados en la introducción de esta tesis.

La matriz de Mueller M es una matriz de 4x4 y tiene una relación directa con la matriz de Jones. Esto si es que el material por el cual paso la onda es un material sin pérdidas. Es decir, corresponde a un material no depolarizante. Esta relación es la siguiente:

$$M = A(T \otimes T^*)\ddot{A}^{-1} \tag{3.1.6}$$

donde $\otimes$ corresponde al producto tensorial o producto de Kronecker, y la matriz A es una matriz de transformación que tiene la siguiente forma:

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & i & -i & 0 \end{pmatrix} \tag{3.1.7}$$

Debido a que esta transformación solo puede ser usada para medios no depolarizantes, debemos tener alguna expresión que nos permita medir cuanto es la depolarización del medio. Esta expresión se puede obtener usando la ecuación (3.1.5) con la matriz de Mueller del material o medio, y usando como onda de entrada una no polarizada obteniendo:

$$S_{out} = \begin{pmatrix} S_{0o} \\ S_{1o} \\ S_{2o} \\ S_{3o} \end{pmatrix} = \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{3.1.8}$$

El grado de polarización del vector de Stokes que representa el estado final de la onda o rayo es llamado polarización escalar y es denotado por la letra **P**. Este valor es calculado usando los valores del vector de Stokes de salida de la ecuación (3.1.8) en la siguiente ecuación:

$$\boldsymbol{P} = \frac{\sqrt{S_{1o}^2 + S_{1o}^2 + S_{2o}^2}}{S_{0o}^2} \tag{3.1.9}$$

Con este valor de **P** podemos saber que tan aproximada es la transición que vamos a hacer en esta tesis entre una matriz de Mueller y una de Jones del material. Esto debido a que vamos a usar matrices experimentales para hacer algunos cálculos de esta tesis.

3.2 Elipsometría como problema de Machine Learning.

La forma más fácil de hacer polarimetría dentro de un laboratorio es obteniendo midiendo distintas polarizaciones de incidencia en un material (Enric Garcia-Caurel, 2013). En el capítulo 1 vimos que mediante un algoritmo de trazado de rayos podemos obtener la matriz de Jones de un medio. Usando la ecuación (3.1.6), siempre y cuando el medio sea no depolarizante o el grado de polarización sea cercano a uno, podemos estimar la matriz de Mueller del material a partir de la matriz de Jones.

Los datos de entrada del algoritmo de trazado de rayos son las constantes ópticas del material, y las salidas del algoritmo son matrices de Jones que podemos transformar en matrices de Mueller, las cuales pueden ser medidas en un laboratorio. Nuestro objetivo es invertir el problema de manera que a partir de una matriz de Mueller, podamos obtener las propiedades ópticas del material como el ángulo del eje óptico con respecto al plano incidente, los índices de refracción del material y el largo del camino óptico.

Supongamos ahora que tenemos el caso de un material plano e isotrópico, como las películas delgadas que se usan comúnmente en investigación o laboratorios industriales. Usando elipsometría de reflexión, la matriz de Mueller de este material es una matriz diagonal donde su primera componente es el coeficiente de Fresnel de reflexión de la polarización paralela y el segundo de la polarización perpendicular al plano de incidencia. Tomando la siguiente forma:

$$\begin{pmatrix} E_x^{out} \\ E_y^{out} \end{pmatrix} = \begin{pmatrix} r_p & 0 \\ 0 & r_s \end{pmatrix} \begin{pmatrix} E_x^{out} \\ E_y^{out} \end{pmatrix} \tag{3.2.1}$$

Para este caso la elipsometría estándar se simplifica a solo dos valores, los ángulos elipsometricos ψ y Δ, estos ángulos usualmente se definen a partir de la siguiente relación

$$\rho = \frac{r_p}{r_s} = \tan\ \psi e^{i\Delta} \tag{3.2.2}$$

donde

$$\tan\ \psi = \frac{|r_p|}{|r_s|}\ \ y\ \Delta = \ \delta_p - \delta_s \tag{3.2.3}$$

Así, $\tan\ \psi$ es la relación de amplitud en la reflexión y Δ es la diferencia entre las fases.

Dentro de las respuestas que se pueden tener al hacer elipsometría podemos encontrar que existen tres propiedades fundamentales, las cuales son la atenuación, retardación y depolarización. Una muestra puede ser puramente atenuador, puramente retardador o puramente depolarizador. Aunque normalmente una muestra desconocida es una mezcla de estas tres propiedades. Debido a que este problema es más complejo que para un material isotrópico plano o algún material que tenga propiedades puras, existen métodos que permiten simplificar este problema usando descomposición de matrices de Mueller, las cuales se pueden clasificar en dos tipos donde la primera es la descomposición por suma y la segunda es una descomposición por producto.

La descomposición por suma más usada es la descomposición de Cloude. La cual se basa en que cualquier matriz de Mueller se puede descomponer en una suma de cuatro matrices no depolarizantes de Mueller, las cuales cada una tiene un peso asociado de la siguiente forma (S.R. Cloude,1996)

$$M = \lambda_1 M_1 + \lambda_2 M_2 + \lambda_3 M_3 + \lambda_4 M_4 \tag{3.2.5}$$

donde cada peso λ_i debe ser un valor positivo.

Otro caso de descomposición por suma es la descomposición de Le Roy-Bréhonnet, la cual es un caso especial de la descomposición de Cloude, ya que esta si contempla depolarización en el sistema. Esta descomposición se basa en la suma de dos matrices donde la primera matriz es una no depolarizante y la segunda es una matriz diagonal depolarizante y este tipo de

descomposición solo es válida cuando se cumple que $\lambda_2 = \lambda_3 = \lambda_4 \neq \lambda_1$, y tiene la siguiente forma (F. Le Roy-Bréhonnet, 1997)

$$M = M_{nd} + M_{id} \tag{3.2.6}$$

Dentro de la descomposición por productos existen tres casos, la primera descomposición es la descomposición hacia adelante e inversa en tres factores la cual tiene la particularidad que se basa en separar en una multiplicación de una matriz puramente atenuadora, otra puramente retardadora y otra puramente depolarizante. Debido al uso de tres matrices para hacer la multiplicación existen seis posibles órdenes para hacer la descomposición donde la más usada es la siguiente (S.Y. Lu, 1996)

$$M = M_P M_R M_D \tag{3.2.6}$$

donde la primera matriz es la depolarizante, la segunda es la retardadora y la tercera la atenuadora.

Otro caso es la descomposición simétrica, la cual se basa en separar una matriz no depolarizante en tres matrices, donde la primera y la última son retardadores lineales, y la segunda es un atenuador retardador lineal con orientación en sus ejes de retardación y atenuación conocidos, de la siguiente forma (R. Ossikovski, 2008)

$$M = M_{LR2} M_{\Delta\psi} M_{LR1} \tag{3.2.7}$$

este método se puede generalizar a matrices depolarizantes en una descomposición, se basa en la descomposición por 5 matrices de la siguiente forma

$$M = M_{D2} M_{R2} M_{d\delta} M_{R1} M_{D1} \tag{3.2.8}$$

Donde $M_{D1}\ y\ M_{D2}$ representan atenuadores genéricos, $M_{R1}\ y\ M_{R2}$ representan retardadores genéricos y $M_{d\delta}$ es un depolarizante diagonal (R. Ossikovski, 2009).

El último caso es la descomposición logarítmica, la cual es una alternativa a la descomposición de productos estándar, esta descomposición se basa en la siguiente ecuación

$$\frac{dM}{dz} = mM \tag{3.2.9}$$

donde la matriz *m* es de 4x4 y contiene las propiedades del medio (R. Ossikovski, 2011).

Para poder extraer información de las propiedades físicas de la muestra se necesitan utilizar métodos indirectos, en este caso la elipsometría como se muestra en la figura 3.2.1. Esto quiere decir que para un conjunto de data experimental dado, se necesita construir un modelo teórico de la muestra que reproduzca tan cercano como sea posible los datos medidos. Esta técnica está limitada al análisis de muestras que no depolarizan la luz, ya que a la mínima depolarización del sistema se produce una superposición incoherente en haces de luz con diferentes estados de polarización (E. Garcia-Caurel, 2013).

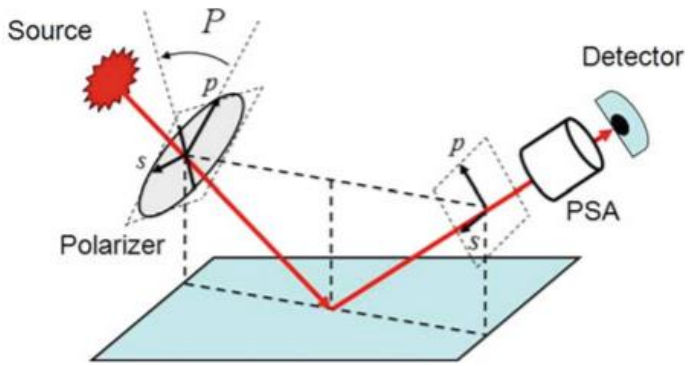

Figura 3.2.1. Esquema general de un experimento de elipsometría de reflexión estándar, donde el PSA es el analizador de estados de polarización.

Debido a esto es que en esta tesis queremos crear un método más general para obtener las propiedades del material mediante elipsometría. También se puede apreciar que para hacer este tipo de inversión existen varías no linealidades dentro del problema, por lo que usar algoritmos de Machine Learning es una opción viable para generalizar el problema. Para hacer esta inversión mediante Machine Learning debemos proponernos este problema como uno de este tipo, por lo que usaremos la metodología de la figura 3.2.3 para llevarlo a cabo.

Para llevar a cabo esta inversión debemos explicar una neural network que se introdujo en la subsección 2.4 llamada concatenate. Esta neural network toma por lo menos dos hidden layers y los junta, concatenándolos en una sola, introduciendo mayor profundidad a la red con una menor cantidad de neuronas. Esta neural network sigue el esquema de la figura 3.2.2.

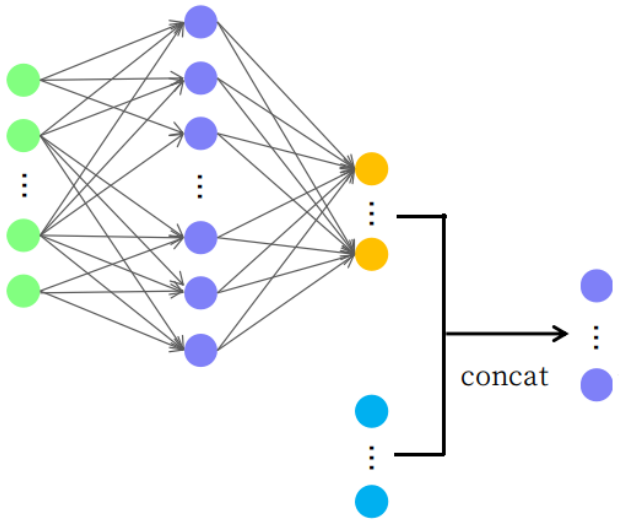

Figura 3.2.2 Esquema estándar de una concatenate neural network, donde las hidden layers en amarillo y en celeste contienen las neuronas de entrada.

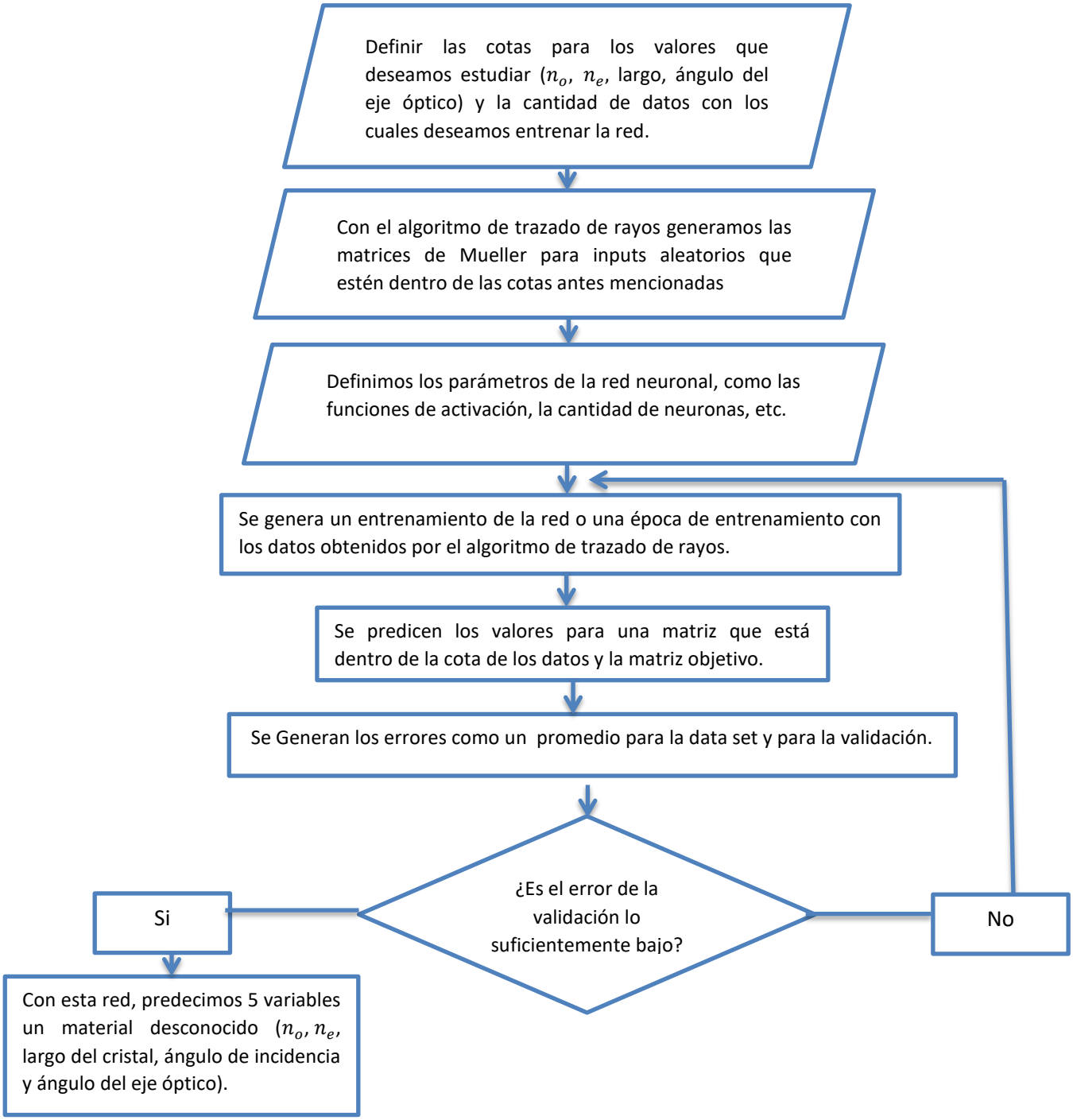

Figura 3.2.1 Diagrama de flujo del algoritmo de inversión usando como base el algoritmo de trazado de rayos y las redes neuronales.

3.3 Limitaciones del algoritmo.

Para resolver el algoritmo de trazado de rayos necesitamos resolver una ecuación (1.5.4) de autovalores para el campo eléctrico. En algunos casos existen ciertos problemas para obtener un autovalor que resuelva esta ecuación matricial, en este caso este problema es una degeneración de los autoestados de la ecuación. Esto significa que, supongamos que tenemos una ecuación vectorial de la siguiente forma

$$A\vec{x} = 0 \tag{3.3.1}$$

donde A es una matriz conocida de dimensiones $n \times n$ y $\vec{x}$ es el vector que queremos calcular de dimensiones $1 \times n$. Para solucionar este problema debemos encontrar los autovalores y autovectores del problema, donde el autovector que corresponda al autovalor cero o más cercano a cero es el que es solución del problema. En algunos casos existen dos o más autovalores soluciones para el problema de autovalores, pero solo uno de estos autovectores es solución por lo que el subespacio de soluciones del problema se encontrara degenerado (E. Merzbacher, 1998). Esto quiere decir que para encontrar el resultado físico del problema debemos probar todos los autovalores que sean posibles soluciones y optar por el que tiene un significado físico.

Como sabemos esta ecuación de autovalores se debe resolver cada vez que la luz atraviese la interfaz birrefringente, esto quiere decir que transmita en un material birrefringente o que refleje en uno, para esta tesis tenemos el caso isotrópico-birrefringente-isotrópico, debido a esto debemos hacer uso de esta ecuación 6 veces, estos casos son los dos campos transmitidos para el caso isotrópico-birrefringente, y cuatro campos reflejados en la interfaz birrefringente-isotrópico.

Debido a que la solución de la ecuación $(1.5.30)$ depende principalmente de la matriz de Fresnel (ecuación (1.5.31)) y de los campos de incidencia, la solución del problema de autovalores y los campos incidentes tienen repercusión directa con los coeficientes de Fresnel, por lo que el problema varía según la polarización incidente como se puede ver en las figuras 3.3.1 y 3.3.2

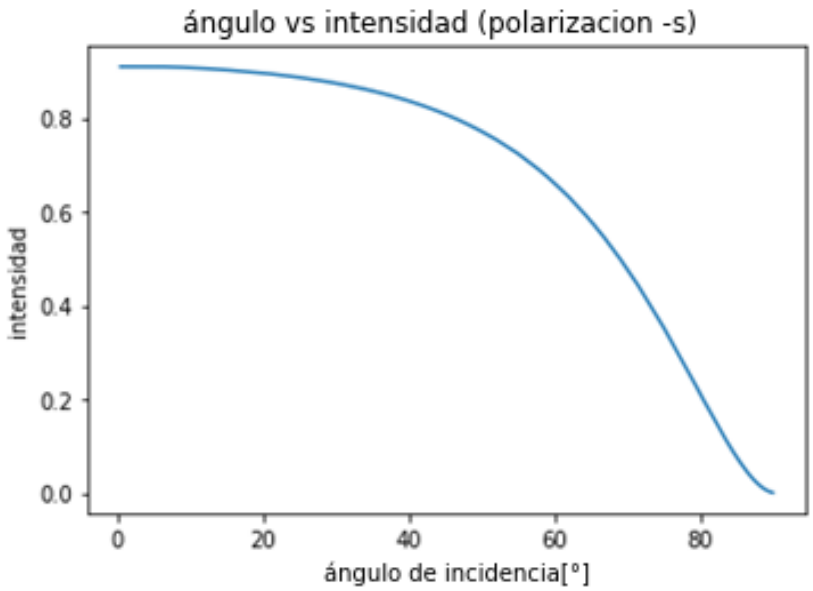

Figura 3.3.1. Intensidad relativa con respecto a la intensidad incidente versus ángulo de incidencia para un cuarzo tipo α, con un ángulo del eje óptico con respecto al plano de incidencia de 43 grados con polarización incidente horizontal (-s).

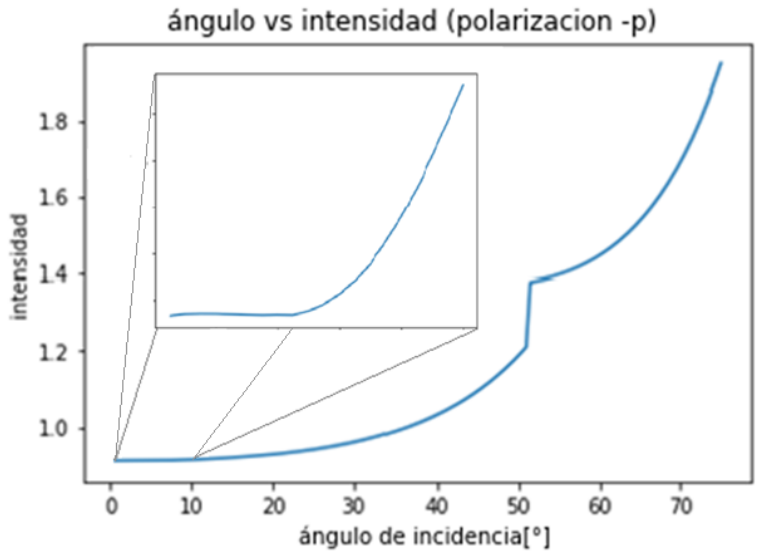

Figura 3.3.2. Intensidad relativa con respecto a la intensidad incidente versus ángulo de incidencia para un cuarzo tipo α, con un ángulo del eje óptico con respecto al plano de incidencia de 43 grados con polarización incidente vertical (-p), con un zoom en el grafico entre los 0 y 10 grados de incidencia.

El comportamiento del gráfico de la figura 3.3.1 (polarización Horizontal) es el comportamiento habitual para la intensidad de un rayo que pasa por un material. Debido a los problemas con de degeneración de autovalores, el comportamiento en el grafico 3.3.2 no es físico ya que la intensidad de salida sobrepasa el valor 1 y el medio no tiene ganancia. Este comportamiento se puede apreciar de mejor forma en el zoom mostrado en el mismo grafico que muestra un comportamiento normal para ángulos bajos de incidencia (entre los 0 y 4 grados) y un comportamiento no físico para ángulos de incidencia mayores. Debido a esto, en esta tesis se decidió trabajar con incidencia normal para todas las polarizaciones de incidencia.

3.4 Caso de estudio A: cuarzo tipo α con matriz de Mueller simulada.

Para validar el algoritmo, lo primero que se hará en esta tesis será intentar reproducir las constantes del cuarzo (Stephen C. McClain, (1992), Algorithms). Con estos resultados podemos tener una estimación de la validez para la inversión usando redes neuronales. Los valores de los índices de refracción del cuarzo que son n_o= 1.53917 y n_e= 1.54811 a 762 nm (Stephen C. McClain, (1992), Algorithms), un ángulo de incidencia igual (θ_i) a cero grados y un ángulo del eje óptico (θ_c) de unos 43 grados y un largo de 6.35 mm.

Usando el algoritmo de trazado de rayos, se obtuvieron matrices de Mueller con las siguientes cotas, en n_o la cota es de 1.50 a 1.59, para n_e la cota es de 1.5 a 1.59, para el largo del cristal la cota es una constante de 6.35 [mm], al igual que en el caso del largo del cristal la cota para el ángulo de incidencia también es una constante que es de 0 grados y para el ángulo del eje óptico la cota es de 40 a 50 grados. Usando estas cotas generamos valores aleatorios uniformes para cada uno de las variables, generando un total de 30.000 matrices para entrenar, ósea unos 480.000 datos en total para entrenar y unas 12.000 matrices para validar (192.000 datos). Luego usando una red neuronal se pueden obtener las constantes del cuarzo como datos de predicción. La red neuronal que vamos a usar para este problema tiene los parámetros y forma mostrados en la figura 3.4.1.

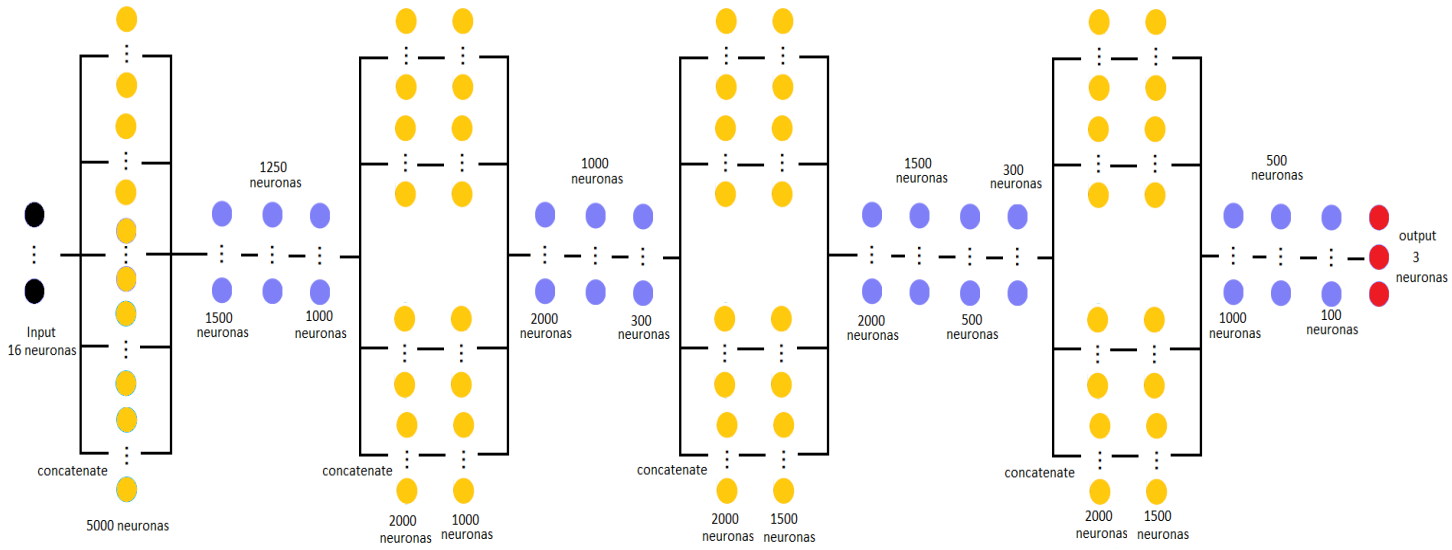

Figura 3.4.1. Esquema de la red neuronal usada para hacer la inversión del cuarzo tipo α, donde cada columna es una hidden layer y las neuronas en negro son las entradas, las en amarillo representan hidden layers usadas en una concatenación, las azules representan típicas Dense hidden layers y las rojas representan neuronas de salida.

Todas las hidden layers tienen un Dropout de 0.3 o 30% excepto por las de salida, ya que equivale a los datos que realmente queremos obtener de la red neuronal, y son 3 en vez de los 5 propuestos en el esquema de la figura 3.2.1 debido a que estos dos valores son constantes. También todos los hidden layers, excepto la entrada y salida de la red neuronal, están normalizados (Batch_Normalization), esto para que la función de optimización converja con mayor facilidad, ya que como se mencionó en el capítulo 2, las redes neuronales suelen trabajar de mejor forma con valores entre 0 y 1. Para terminar la red neuronal tenemos que la función de optimización es "Adam" con un Learning rate de 0.1 usando un decaimiento por época de un 0.01, compilándolo con una función de perdida "logcosh" (cuya ecuación es la 3.4.1 y su grafico se muestra en la figura 3.4.2), la métrica a seguir es de error cuadrático medio y con un batch size de 400 datos por actualización de la red, el optimizador es cambiado por "SGD" después de la época 20, el cual tiene un Learning rate de 0.05, un decaimiento por época de 0.1 y un momentum de 0.8, esto debido a que SGD funciona de mejor forma para la convergencia de un problema.

$$L(y, y^p) = \sum_{i=1}^{n} \log\left(\cosh\left(y_i^p - y_i\right)\right) \tag{3.3.1}$$

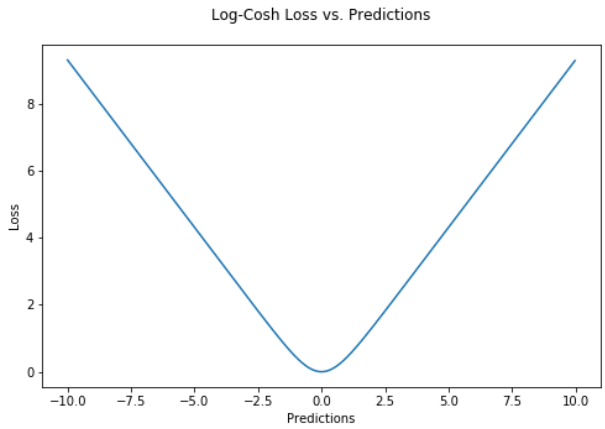

Figura 3.4.2. Gráfico de la función "logcosh" usada como función de pérdida en la red neuronal aplicada en el algoritmo de inversión para el cuarzo tipo α. Una de las ventajas de esta función es doblemente diferenciable.

Cabe mencionar que todos estos valores no fueron puestos de forma aleatoria sino que fueron probados en conjunto y por separado para poder tener una mejor predicción con la red neuronal planteada para este problema. Si el problema varia entonces la red debe varia.

Esta red neuronal tiene 86.582.573 parámetros, donde 86.457.913 son parámetros posibles para entrenar y 124.660 son parámetros no posibles para entrenar dentro de la red, estos datos no posibles por entrenar son los datos puestos manualmente dentro de la red como la cantidad de hidden layers o la cantidad de neuronas que tenga cada hidden layer. Esta red fue entrenada con 30.000 datos y con 12.000 datos de validación (porcentaje de un 40% de datos usados para la validación), usando 662 épocas demorando un aproximado de 45 segundos por cada época, con un sistema de corte del entrenamiento basado en la perdida de datos de validación el cual fue de 0.4, se usó un batch size de 1.000 datos por actualización de la red, obteniendo los siguientes resultados.

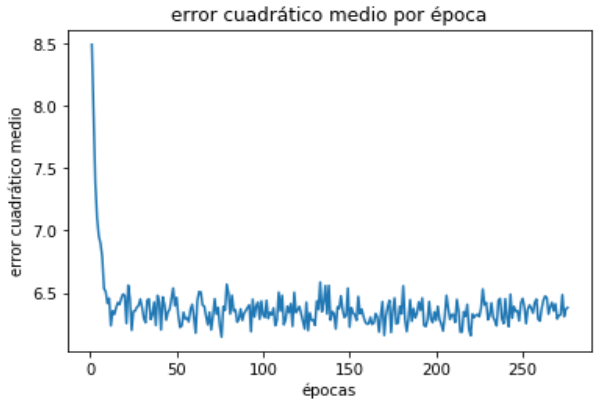

Figura 3.4.3. Gráfico del error cuadrático medio por época para los datos de entrenamiento obtenidos para la red neuronal mencionada anteriormente para el cuarzo tipo α.

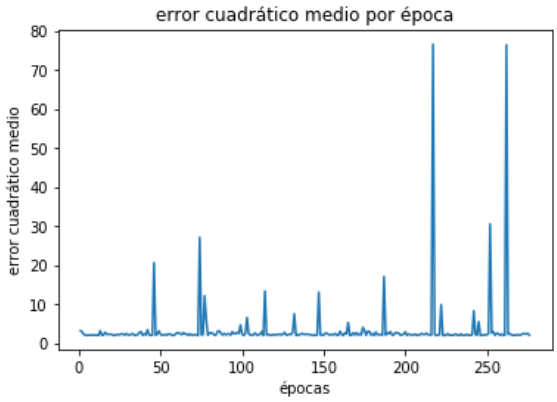

Figura 3.4.4. Gráfico del error cuadrático medio por época para los datos de validación obtenidos para el entrenamiento de la red neuronal mencionada anteriormente para el cuarzo tipo α.

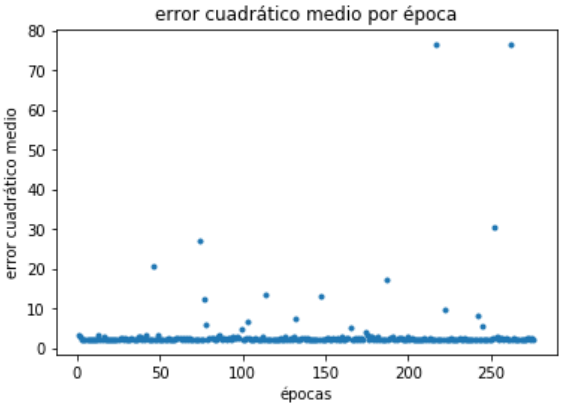

Figura 3.4.5. Gráfico del error cuadrático medio por época para los datos de validación obtenidos para el entrenamiento de la red neuronal mencionada anteriormente para el cuarzo tipo α, en este grafico cada dato obtenido para el erro cuadrático medio es un punto.

Como se observa en la figura 3.4.3 el error cuadrático medio llego a una meseta al igual que para los datos de validación por lo que se asume que la red puede llegar a mejorar pero esta pronto al punto mínimo de error. Esta red neuronal puede mejorar si es que se deja entrenar por más tiempo o aumentando la cantidad de datos de entrenamiento y validación, o cambiando el número de neuronas o probando algún otro tipo de configuración.

Para el la red neuronal antes mencionada a las 284 épocas de entrenamiento, con un aproximado de 40[s] por época, se obtuvo que para los datos de entrenamiento se tiene una pérdida de 0.3266 y un error cuadrático medio de 1.2860, y para los datos de validación se tiene una pérdida de 0.14923 y un error cuadrático medio de 0.3896. Cabe mencionar que la precisión de la red es 1 tanto para la validación como para los datos de entrenamiento, esto ocurre desde la tercera época en adelante, por lo que la precisión no pasa a ser un dato importante para la red.

Cabe mencionar que la configuración usada en este caso no fue al azar, ya que se probaron distintas configuraciones donde esta red fue la que obtuvo mejores resultados de convergencia, obteniendo las siguientes predicciones

Matriz obtenida por datos tabulados	Valor tabulado	Valor predicho	Error relativo porcentual
n_o	1.5391700	1.5669432	2.44%
n_e	1.5481100	1.5767944	2.19%
θ_c	43.0000°	43.941322°	1.21%

Tabla 3.4.1. Tabla de comparación de datos tabulados con los predichos para el caso de una matriz de Mueller obtenida por el algoritmo de trazado de rayos para un cristal de cuarzo tipo α con n_o= 1.53917, n_e= 1.54811, θ_i=0°, θ_c= 43°, un largo de 6.35 mm y con una longitud de onda de 762 nm.

También se hizo la prueba con una matriz de testeo, la cual fue usada para ver si la red neuronal estaba haciendo overfitting o no, la cual consiste en una matriz de Mueller de puros valores dos. Para dicha matriz los valores predichos son los siguientes:

Matriz de testeo	Valores predichos
n_o	109.9883
n_e	101.225555
θ_c	347.5287°

Tabla 3.4.2. Tabla de valores predichos para una matriz de Mueller de testeo en la cual todas sus componentes son ceros, estos datos sirven para comprobar si existe overfitting en la red.

Como podemos ver para el caso de la matriz de testeo los tres valores están fuera de las cotas usadas para entrenar la red, por lo que nos dice que la red no logra hacer una regresión apropiada para estos datos. Además podemos ver que el error entre la predicción y el valor tabulado tanto para los índices de refracción como para el ángulo de incidencia están dentro del error cuadrático medio obtenido por la red neuronal y ambos valores relativos porcentuales son menores que 3%.

3.5 Caso de estudio B: BBO usando una matriz experimental.

Usando la misma red neuronal propuesta para la regresión en el caso del cuarzo tipo α, se quiere entrenar una red neuronal que tenga la capacidad no solamente de hacer una predicción prudente para el caso de una matriz obtenida por medio del algoritmo de trazado de rayos, sino que se quiere obtener una predicción de una matriz obtenida experimentalmente usando un método parecido a los mencionados en la introducción de esta tesis, la cual es de un cristal BBO, cuya matriz de Mueller fue obtenida en los por

colaboradores. Sabemos que para el BBO con una longitud de onda de 853[nm] tenemos que el $n_o = 1.6589$ y el $n_e = 1.5446$ (G. Tamošauskas, 2018), con un ángulo de incidencia de 49°. Para este caso las cotas para los datos de entrenamiento son los siguientes: el índice de refracción ordinario varía entre [1.6-1.7], el índice de refracción extraordinario está en la cota [1.5-1.6] y el ángulo del eje óptico varía de 27 a 31 grados, con un ángulo de incidencia de 0 grados (incidencia normal) y un largo de cristal de 3 [mm], usando una longitud de onda de 853[nm], además debido a que el BBO es un material no quiral su tensor gyrotropico es una matriz de puros ceros, pero debido a limitaciones numéricas del problema usamos la matriz diagonal de la ecuación (1.5.7) con $g_o = 10^{-11}$ y $g_e = 0$ (estos valores sirven para estabilizar el algoritmo). Como se puede ver los dos índices de refracción tabulados están dentro de las cotas con las cuales se obtuvieron los datos de entrenamiento. En este caso, el algoritmo debe parar cuando el error cuadrático medio sea menor que 0.4. Con esta red neuronal se obtuvieron los siguientes resultados

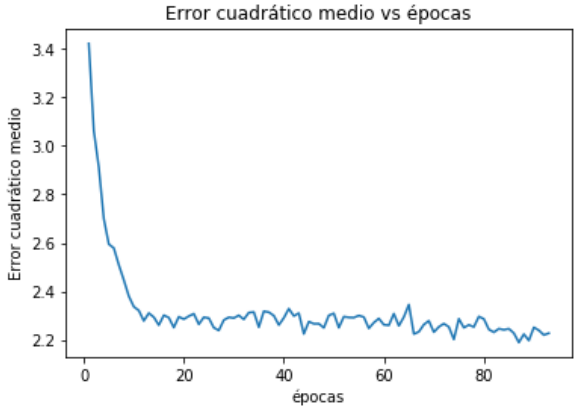

Figura 3.5.1. Gráfico del error cuadrático medio por época para los datos de entrenamiento obtenidos para el entrenamiento de la red neuronal mencionada anteriormente para el BBO.

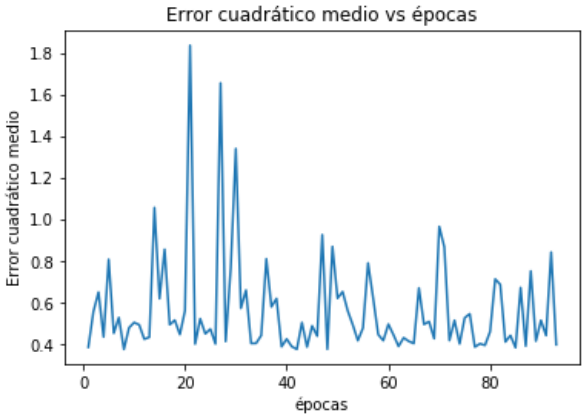

Figura 3.5.2. Gráfico del error cuadrático medio por época para los datos de validación obtenidos para el entrenamiento de la red neuronal mencionada anteriormente para el BBO.

Para la misma red neuronal usada en el caso de A (figura 3.4.1), con un entrenamiento de 94 épocas y un aproximado de 62[s] por época, se obtuvo que para los datos de entrenamiento se tiene una pérdida de 0.5756 y un error cuadrático medio de 2.2275, y para los datos de validación se tiene una pérdida de 0.1551 y un error cuadrático medio de 0.3994. Al igual que en el caso del cuarzo la precisión de la red es de 1, debido a esto la precisión de la red no es un dato que cuantifique de forma fehaciente la calidad de la red neuronal, este dato se dejó de lado.

Al igual que en caso del cuarzo el grafico de la figura 3.5.1 se puede apreciar que el error cuadrático medio llego a una especie de meseta en la cual la minimización de este error cuesta cada vez más recursos, ósea cuesta mayor cantidad de iteraciones o épocas, la opción de mejorar esto es cambiar las variables dentro de la red o trabajar con una cantidad mayor de datos. Aunque para este problema el resultado es bastante cercano a lo que se quiere llegar con errores de 2% como máximo para una matriz obtenida con el algoritmo de ray tracing usando los datos tabulados, como se puede apreciar en los siguientes resultados

Matriz obtenida por datos tabulados	Valor tabulado	Valor predicho	Error relativo porcentual
n_o	1.65890	1.6575224	0.083%
n_e	1.54460	1.5571532	0.074%
θ_c	29.0000°	29.579094°	1.997%

Tabla 3.5.1. Tabla de comparación de datos tabulados con los predichos para el caso de una matriz de Mueller obtenida por el algoritmo de trazado de rayos para un cristal de BBO con $n_o = 1.65890$, n_e=1.54460, θ_i=0°, θ_c= 29°, un largo de 3 mm y con una longitud de onda de 853 *nm.*

Como el principal enfoque de esta tesis es generar un nuevo método de caracterización de materiales a través de polarimetría, se quiere saber si el método sirve para obtener las propiedades de un material cuya matriz de Mueller se obtuvo experimentalmente usando métodos de polarimetría. Para esto usamos la matriz obtenida por los colaboradores de un cristal de BBO con incidencia normal cuyo ángulo de eje óptico se estima que es de 49 [°], cuya matriz de Mueller es la siguiente

$$M_{bbo} = \begin{pmatrix} 1 & 0.0017 & -0.0282 & -0.0016 \\ 0.0074 & 0.9809 & 0.1450 & -0.1227 \\ 0.0008 & 0.1150 & 0.0116 & 0.8405 \\ 0.0280 & 0.1280 & -0.8389 & -0.0057 \end{pmatrix} \tag{3.5.1}$$

Esta matriz tiene un grado de polarización igual a 0.9018478622620706, ósea que tiene un grado de depolarización de un 10% aproximadamente, el cual está calculado usando la ecuación (3.1.9). Esto quiere decir que aparte del error estimado por la red neuronal, tenemos un error agregado por la depolarización de sistema, ya que la transformación de matriz de

Jones a matriz de Mueller solo puede hacerse con materiales no depolarizantes o con materiales con polarización igual a 1.

Matriz BBO obtenida experimentalmente (ecuación 3.5.1)	Valor tabulado	Valor predicho	Error relativo porcentual
n_o	1.65890	1.6874924	1.723%
n_e	1.54460	1.5856276	1.883%
θ_c	29.0000°	29.614323°	2.118%

Tabla 3.5.2. Tabla de comparación de datos tabulados con los predichos para el caso de una matriz de Mueller obtenida experimentalmente para un cristal de BBO (ecuación 3.5.1) con $n_o = 1.65890$, n_e=1.54460, θ_i=0°, θ_c= 29°, *un largo de 3 mm y con una longitud de onda de* 853 *nm.*

Debido a que los errores relativos porcentuales de la tabla 3.5.2 son más grandes que para el caso de la matriz obtenida por el algoritmo de ray tracing, podemos asociarlos a la depolarización del sistema.

Al igual que en el caso del caso del cuarzo se generó una matriz de Mueller cuyas componentes son puros valores dos, esto se hace para ver si la red produce overfitting o no. Para esta matriz se predijo lo siguiente

Matriz de testeo	Valores predichos
n_o	7.4822151
n_e	7.0209793
θ_c	135.87839°

Tabla 3.5.3. Tabla de valores predichos para una matriz de Mueller de testeo en la cual todas sus componentes son dos, estos datos sirven para comprobar si existe overfitting en la red.

Los datos de la tabla 3.5.3 nos aseguran que la regresión no-lineal que está haciendo la red neuronal es óptima, ósea no existe overfitting, ya que como vimos en el caso de una interpolación de la función coseno (capitulo 2), los valores de predicción son acertados solo para hacer una interpolación que este dentro de los datos que fueron usados para entrenar a la red.

Conclusiones.

En esta tesis se desarrolló un algoritmo de trazado de rayos utilizando las ecuaciones de Maxwell y las ecuaciones de Fresnel para estudiar la refracción que se produce en materiales birrefringentes en incidencia y en transmisión por el material. Con el algoritmo se obtuvieron los campos eléctricos y magnéticos de salida del material. Comparando estos campos con los de entrada obtuvimos una matriz de Jones, con la cual podemos obtener una matriz de Mueller para medios no depolarizantes. Usando este algoritmo se entrenó una red neuronal, con el objetivo de resolver el problema de regresión en elipsometría, es decir, a partir de la matriz de Mueller del material obtener las constantes ópticas del material.

Para el cuarzo tipo α se trabajó con una red neuronal concatenada, para poder aumentar la profundidad de la red sin aumentar la cantidad de neuronas, como se vio en el capítulo 3. Las matrices de entrenamiento y validación fueron obtenidas en incidencia normal, y se logró resolver el problema de inversión en cuestión bajo nuestra aproximación numérica, podemos obtener una red neuronal para hacer regresión de datos polarimetricos en materiales no depolarizantes.

Usando la misma red neuronal para el caso del cuarzo, se estudió el material no quiral BBO. Usamos una matriz obtenida experimentalmente por colaboradores a través de polarimetría de transmisión y comparamos las predicciones de constantes ópticas obtenidas a partir de matrices de una Mueller experimental y numéricamente del algoritmo de trazado de rayos en el caso de matrices no obtenidas por este algoritmo.

El método desarrollado en esta tesis permite obtener con una precisión mayor a un 3% máximo de error, los índices de refracción del BBO dentro de un rango de valores de entrenamiento restringido. Por lo tanto ampliando el problema a una mayor cantidad de longitud de ondas, podemos obtener una ecuación de Sellmeier para el material en cuestión, permitiendo así obtener la longitud de onda en la cual las propiedades ópticas de segundo orden del material sean óptimas.

Los desafíos futuros que deja esta tesis son extender el algoritmo de trazado de rayos más allá de incidencia normal para todo tipo de polarización incidente, esto haría más fácil la comparación con las matrices de Mueller experimentales. Otro desafío es tener más variables obtenibles para la red como por ejemplo la longitud de onda del material, el largo del cristal y el ángulo de incidencia mencionado anteriormente. Por ultimo escalar el problema de forma que las cotas de parámetros para entrenar la red sean más amplias, así poder abarcar de mejor forma los índices de refracción de una mayor cantidad de materiales. Con el objetivo de generalizar el método de extracción de constantes opticas.

Referencias Bibliográficas

Industrial Aplications of Lasers, Jhon F. Ready, 1978

G. Q. Zhou, J. Xu, X. D. Chen, H. Y. Zhong, S. T. Wang, K. Xu, P. Z. Deng and F. X. Gan, J. Cryst. Growth, 191 (1998), 517−519.

W. Sellmeier, Zur Erklärung der abnormen Farbenfolge im Spectrum einiger Substanzen, *Annalen der Physik und Chemie* 219, 272-282 (1871).

Appel, R., Dyer, C. D., & Lockwood, J. N. (2002). Design of a broadband UV–visible alpha-barium borate polarizer. Applied Optics, 41(13), 2474.

G. Tamosauskas, G. Beresnevicius, D.Gadonas, A. Dubietis. Transmittance and phase matching of BBO crystal in the 3–5 µm range and its application for the characterization of mid-infrared laser pulses (2018)

Xinming Wan, H. L. W. Chan, C. L. Choy, Xiangyong Zhao, and Haosu Luo. Optical properties of (1 - x) Pb (Mg 1／3 Nb 2／3) O 3 - x PbTiO 3 single crystals studied by spectroscopic ellipsometry. Journal of Applied Physics 96, 1387 (2004); doi: 10.1063/1.1767287.

R. Paschotta, the *Encyclopedia of Laser Physics and Technology*, 1. edition October 2008, Wiley-VCH, ISBN 978-3-527-40828-3 P. A.

Franken *et al.*, “Generation of optical harmonics”, Phys. Rev. Lett. 7 (4), 118 (1961) (first report of second-harmonic generation, not yet phase-matched)

A. Ashkin, G. D. Boyd, and J. M. Dziedzic, “Resonant optical second harmonic generation and mixing”, IEEE J. Quantum Electron. 2 (6), 109 (1966)

Robert W. Boyd. Nonlinear Optics. Second Edition. 2003

R. Ulrich *et al.*, “Bending-induced birefringence in single-mode fibers”, Opt. Lett. 5 (6), 273 (1980)

M. Bass, Eric W . Van Stryland, David R . Williams, William L . Wolfe. Handbook Of Optics, Second edition, Volume 1, 1995

R . M . A . Azzam and N . M . Bashara , Ellipsometry and Polarized Light , North-Holland , Amsterdam , 1987.

R . M . A . Azzam and A . G . Lopez , J . Opt . Soc . Am . A 6 : 1513 – 1521 (1989)

Nonlinear Optics Theory, Numerical Modeling, and Applications. Partha P. Banerjee, 2003

Jackson John David. Clasical electrodinamics. Third Edition. 1999.

Robert W. Boyd. Nonlinear Optics. Second Edition. 2003

Robin H. A. Ras, Robert A. Schoonheydt, and Cliff T. Johnston. Relation between s-Polarized and p-Polarized Internal Reflection Spectra: Application for the Spectral Resolution of Perpendicular Vibrational Modes. J. Phys. Chem. A 2007, 111, 8787-8791.

Fundamentos de Óptica para Ingeniería Informática. A. Beléndez, 1996

Fundamentals of photonics. Bahaa E. A. Saleh, Malvin Carl Teich, 1991

Handbook of Nonlinear Optical Crystals. V.G. Dmitriev, G. G. Gurzadyan, D.N Nikogosyan, 1999

"Theories of optical rotatory power". E.U. Condor, 1937

Aberration Of Optical Systems, W.T. Welford, 1986.

Polarization ray tracing in birefringent media. J.D. Trolinger, Jr., R. A. Chipman and D.K Wilson (1991)

Polarization ray tracing in anisotropic optically active media I. Algorithms. Stephen C. McClain, Lloyd W. Hillman, and Russell A. Chipman. 1993

Polarization ray tracing in anisotropic optically active media II. Theory and physics. Stephen C. McClain, Lloyd W. Hillman, and Russell A. Chipman. 1993

Introduction to Machine Learning. Nilsson N.J. 1996

Ethem Alpaydin. Introduction to Machine Learning, 2004

Tom M. Mitchell. Machine Learning. 1997

Chollet, Francois and others. Keras documentation, 2005

Gintaras Tamošauskas, Gvidas Beresnevičius, Darius Gadonas, and Audrius Dubietis. Transmittance and phase matching of BBO crystal in the 3−5 μm range and its application for the characterization of mid-infrared laser pulses. 2018.

Enric Garcia-Caurel, Razvigor Ossikovski, Martin Foldyna, Angelo Pierangelo, Bernard Drévillon, Antonello De Martino. Advanced Mueller Ellipsometry Instrumentation and Data Analysis. 2013.

R. Ossikovski, Interpretation of nondepolarizing Mueller matrices based on singular-value decomposition. J. Opt. Soc. Am. A 25, 473–482 (2008). doi:JOSAA.25.000473

S.R. Cloude, E. Pottier, IEEE Trans. GRS 34, 498 (1996)

S.Y. Lu, R.A. Chipman, Interpretation of Mueller matrices based on polar decomposition. J. Opt. Soc. Am. A 13, 1106–1113 (1996). doi:10.1364/JOSAA.13.001106

F. Le Roy-Bréhonnet, B. Le Jeune, Utilization of Mueller matrix formalism to obtain optical targets depolarization and polarization properties. Prog. Quantum Electron. 21, 109–151 (1997). doi:10.1016/S0079-6727(97)84687-3

R. Ossikovski, Analysis of depolarizing Mueller matrices through a symmetric decomposition. J. Opt. Soc. Am. A 26, 1109–1118 (2009)

R. Ossikovski, Differential matrix formalism for depolarizing anisotropic media. Opt. Lett. 36, 2330–2332 (2011). doi:10.1364/OL.36.002330

Merzbacher, Eugen (1998). *Quantum Mechanics* (3rd ed.). New York: John Wiley. ISBN 0471887021.